Supporting Information for "**Simultaneous Confidence Corridors for Mean Functions in Functional Data Analysis of Imaging Data**" by Yueying Wang, Guannan Wang, Li Wang and Todd Ogden

In this Supporting Information, we describe the triangulation selection algorithm in Subsection A.1. In Subsection A.2, we include more results from simulation studies. In Subsection A.3, we give more description of the ADNI data analyzed in Section 6 in the main article. Section B includes detailed proofs of the theoretical results in the main article.

Web Appendix A Implementation Algorithm and More Numerical Studies

A.1 Algorithm for Triangulation Selection

In this section, we describe the algorithm for selecting the triangulation for bivariate spline smoothing. As proposed in Section 4.2 in the main paper, we choose the triangulation $\triangle_\mu$ based on leave-images-out k-fold cross-validation (CV), and choose the triangulation $\triangle_\eta$ by minimizing a bootstrap estimator of the coverage error of the SCCs. Details of the algorithm for the one-sample case are given in Algorithm A1 below.

A.2 More Results from Simulation Studies

In this section, we present more simulation results from Sections 5.1 and 5.2 in the main paper. For the simulation example presented in Section 5.1, Figures A1–A3 present the 99% SCCs for the quadratic mean function based on sample size $n = 50$, 100 and 200. Figures A4–A6 present the 99% SCCs for the exponential, cubic and sine mean functions with $n = 50$, respectively. Table A1 summarizes the estimated coverage rate of the SCCs based on 1000 replications for $N = 3682$.

Input : Images $\{Y_{ij}\}_{j=1,i=1}^{N,n}$.

Output: Triangulations $\triangle_\mu$ and $\triangle_\eta$.

Step 1. Selecting $\triangle_\mu$ and estimating $\mu(\boldsymbol{z})$. Based on $\{Y_{ij}\}_{j=1,i=1}^{N,n}$, select $\triangle_\mu$ via the leave-images-out k-fold CV, and obtain $\widehat{\mu}(\boldsymbol{z})$ using the BPS method. Define $\widehat{R}_{ij} = Y_{ij} - \widehat{\mu}(\boldsymbol{z}_j)$.

Step 2. Selecting $\triangle_\eta$ from a set of triangulations $\{\triangle_\eta^q, q \in \mathcal{Q}\}$.

foreach $q \in \mathcal{Q}$ **do**

(i) For $i = 1, \ldots, n$, estimate $\widehat{\eta}_i(\boldsymbol{z})$ by smoothing $\widehat{R}_{ij}$ via the bivariate spline smoothing method based on triangulation $\triangle_\eta^q$, and let $\widehat{\varepsilon}_{ij} = \widehat{R}_{ij} - \widehat{\eta}_i(\boldsymbol{z}_j)$.

(ii) Generate an independent random sample $\delta_i^{(b)}$ and $\delta_{ij}^{(b)}$ from $\{-1, 1\}$ with probability 0.5 each, and define $Y_{ij}^{*(b)} = \widehat{\mu}(\boldsymbol{z}_j) + \delta_i^{(b)}\widehat{\eta}_i(\boldsymbol{z}_j) + \delta_{ij}^{(b)}\widehat{\varepsilon}_{ij}$.

(iii) Based on $\left\{Y_{ij}^{*(b)}\right\}_{j=1,i=1}^{N,n}$, obtain the estimators of the mean and covariance functions $\widehat{\mu}^{*(b)}$ and $\widehat{G}_\eta^{*(b)}$ using $\triangle_\mu$ and $\triangle_\eta^q$, respectively.

(iv) For any fixed $\alpha \in (0, 1)$, construct $100(1-\alpha)\%$ SCCs for resampled data $\left\{Y_{ij}^{*(b)}\right\}_{j=1,i=1}^{N,n}$: $\mathcal{B}^{*(b)}(\alpha),\ b = 1, \ldots, B$,

$\mathcal{B}^{*(b)}(\alpha) = \widehat{\mu}^{*(b)}(\boldsymbol{z}) \pm n^{-1/2} q_{1-\alpha}^{*(b)} \widehat{G}_\eta^{*(b)}(\boldsymbol{z}, \boldsymbol{z})^{1/2}$.

end

Select $\triangle_\eta$ by minimizing the objective function

$$\min_{q \in \mathcal{Q}} \int_{\alpha-\delta}^{\alpha+\delta} \left\{ \frac{1}{B} \sum_{b=1}^{B} I(\widehat{\mu} \in \mathcal{B}^{*(b)}(\alpha); \triangle_\eta^q) - (1-\alpha) \right\}^2 d\alpha,$$

for some constant $0 < \delta < \alpha$, which is taken to be 0.005 in our simulation studies.

Algorithm A1: Triangulation selection.

Table A2 provides the type I error and the empirical power of the two-sample test presented in Section 5.2 in the main paper.

To illustrate the benefits of our method, we conduct the following simulation study to compare the proposed SCC with the traditional multiple testing with Bonferroni correction and the cluster threshold-based method (Poldrack et al., 2011). Similar as in Sections 5.1 in the main paper, we generate the images from the following model:

$$Y_{ij} = \mu(\boldsymbol{z}_j) + \sum_{k=1}^{\kappa} \sqrt{\lambda_k} \xi_{ij} \psi_k(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij},\ \boldsymbol{z}_j \in \Omega \subset [0,1]^2.$$

For comparison, we consider the following mean function, which is similar as the exponential function in Example 1 in Section 5.1 in the main paper:

$$\mu(\boldsymbol{z}) = \begin{cases} \exp\left[-30\left\{(z_1-0.5)^2 + (z_2-0.5)^2\right\}\right], & (z_1-0.5)^2 + (z_2-0.5)^2 \le 0.10 \\ 0, & (z_1-0.5)^2 + (z_2-0.5)^2 > 0.10 \end{cases},$$

and the corresponding images are shown in Figure A7. To simulate the within-image dependence, we generate $\xi_{ik} \overset{\text{i.i.d}}{\sim} N(0,1)$ for $i = 1, \ldots, n,\ k = 1, 2$, and orthonormal basis functions $\psi_1(\boldsymbol{z}) = 0.988\sin(\pi z_1) + 0.5,\ \psi_2(\boldsymbol{z}) = 2.157\cos(\pi z_2) - 0.084$. For the eigenvalues, we set $\lambda_1 = 0.2$, $\lambda_2 = 0.05$. We consider $n = 100,\ 200$ and for each image, the number of pixels is set to be the same as in typical brain imaging which is $N = 79 \times 95 = 7505$.

Based on these images, we are interested in testing $H_0:\ \mu(\boldsymbol{z}_j) = 0,\ \boldsymbol{z}_j \in \Omega, j = 1, \ldots, N$, at significance level $\alpha = 0.05$. For the cluster approach, the threshold is usually set by the practitioner's experience and prior knowledge. In this example, we consider three thresholds: 0.1, 0.05 and 0.01, as suggested in Poldrack et al. (2011). For comparison, we consider the following criteria:

- False Positive Rate (FPR): the proportion of pixels within the domain which are discovered incorrectly as positive (significantly different from zero);

- False Negative Rate (FNR): the proportion of pixels within the domain which are discovered incorrectly as negative (not significantly different from zero);

- False Discovery Rate (FDR): the proportion of detected pixels that are false positives.

Table A3 summarizes all results based on 100 replications. Figure A8 shows the discovery of the true signal via different methods for a typical replication with $n = 200$. Based on Table A3 and Figure A8, it is obvious that the pixel-wise inference with Bonferroni correction is very conservative. The FPRs and FDRs of the Bonferroni correction are very close to zero, while the FNRs are very high, even greater than 30%. Although the FPRs and FDRs for the proposed SCC are above zero, they are still very small, usually less than 1%. Meanwhile, the FNRs for the proposed SCC are much smaller than the Bonferroni correction. In addition, one sees that the cluster threshold-based method heavily depends on the choice of threshold. When using 0.01 as the threshold instead of 0.1, the FPR dramatically decreases while the FNR considerably increases. For $n = 200$, the FPR and FDR of the SCC are both smaller than those of the Cluster-threshold method. From Figure A8, we can also see that our method aims at detecting contiguous groups of active pixels because it is able to account for the spatial dependence within data.

A.3 Additional Information of ADNI Data

The ADNI dataset analyzied in Section 6 in the main paper consists of 112 subjects with normal cognitive functions (control group; CON), 213 subjects with mild cognitive impairment (MCI), and 122 subjects who have been diagnosed with Alzheimer's Disease (AD). As described in the data analysis in Section 6 in the main paper, we stratify the data according to sex and age, and the breakdowns are given in Table A1.

Table A1: Two way table of diagnosis vs. gender and age group.

		Diagnosis			Total
		CON	MCI	AD	
Gender	Female	42	77	50	169
	Male	70	136	72	278
Age	Age ≤ 75	54	107	60	221
	Age > 75	58	106	62	226
Total		112	213	122	447

Web Appendix B Technical Assumptions and Proofs

In the following, we use c, C, c_1, c_2, C_1, C_2, etc. as generic constants, which may be different even in the same line. For any sequence a_n and b_n, we write $a_n \asymp b_n$ if there exist two positive constants c_1, c_2 such that $c_1|a_n| \leq |b_n| \leq c_2|a_n|$, for all $n \geq 1$. For a real valued vector $\boldsymbol{a}$, denote $\|\boldsymbol{a}\|$ its Euclidean norm. For a matrix $\mathbf{A} = (a_{ij})$, denote $\|\mathbf{A}\|_\infty = \max_{i,j} |a_{ij}|$. For any positive definite matrix $\mathbf{A}$, let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ be the smallest and largest eigenvalues of $\mathbf{A}$.

B.1 Assumptions

Given a triangle $T \in \triangle$, let ϱ_T be the radius of the largest disk which can be inscribed in T. Define the shape parameter of T as the ratio $\pi_T = |T|/\varrho_T$. When π_T is small, the triangles are relatively uniform in the sense that all angles of triangles in the triangulation $\triangle$ are roughly the same. Next, we introduce some technical assumptions.

(A1) The bivariate function $\mu(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega) = \{g : |g|_{k,\infty,\Omega} < \infty, 0 \leq k \leq d+1\}$ for an integer $d \geq 1$.

(A2) For any $k \geq 1$, ξ_{ik}'s are i.i.d. random variables with mean 0, variance 1 and $E\,|\xi_{ik}|^{4+\delta_1} < +\infty$

for some constant $\delta_1 > 0$. For any $i = 1, \ldots, n$, $j = 1, \ldots, N$, ε_{ij}'s are i.i.d with mean 0, variance 1, and $E\left|\varepsilon_{ij}\right|^{4+\delta_2} < +\infty$ for some constant $\delta_2 > 0$.

(A3) The function $\sigma \in \mathcal{C}^{(1)}(\Omega)$ with $0 < c_\sigma \leq \sigma(\boldsymbol{z}) \leq C_\sigma \leq \infty$ for any $\boldsymbol{z} \in \Omega$; for any k, $\psi_k \in \mathcal{C}^{(1)}(\Omega)$ and the variance function $0 < c_G \leq G_\eta(\boldsymbol{z}, \boldsymbol{z}) \leq C_G \leq \infty$, for any $\boldsymbol{z} \in \Omega$.

(A4) The triangulation is π-quasi-uniform, that is, there exists a positive constant π such that $(\min_{T\in\triangle} \varrho_T)^{-1}|\triangle| \leq \pi$.

(A5) As $N \to \infty$, $n \to \infty$, $N^{-1}n^{1/(d+1)}\log(n) \to 0$, the triangulation size satisfies that $N^{-1}\log(n) \ll |\triangle|^2 \ll \min\{n^{(2+\delta_2)/(4+\delta_2)}N^{-1}\log^{-1}(n), n^{-1/(d+1)}\}$, and the smoothing penalty parameter ρ_n satisfies $\rho_n = o\{\min(n^{1/2}N|\triangle|^3, nN^{3/2}|\triangle|^6, nN|\triangle|^5)\}$.

(A6) For $k \in \{1, \ldots, \kappa\}$ and a nonnegative integer s, $\phi_k(\boldsymbol{z}) \in \mathcal{W}^{s+1,\infty}(\Omega)$, $\sum_{k=1}^{\kappa} \|\phi_k\|_\infty < \infty$. $\frac{\rho_n}{nN|\triangle|^3}\sum_{k=1}^{\kappa_n}\|\phi_k\|_{2,\infty} = o(1)$, $\left(1 + \frac{\rho_n}{nN|\triangle|^5}\right)\sum_{k=1}^{\kappa_n}|\triangle|^{s+1}\|\phi_k\|_{s+1,\infty} = o(1)$ for a sequence $\{\kappa_n\}_{n=1}^{\infty}$ of increasing integers, with $\lim_{n\to\infty}\kappa_n = \kappa$, as $n \to \infty$. Meanwhile, $\sum_{k=\kappa_n+1}^{\kappa}\|\phi_k\|_\infty = o(1)$. The number κ of nonzero eigenvalues is finite or κ is infinite.

(A7) As $N \to \infty$, $n \to \infty$, for some $0 < \delta_3 < 1$, $N^{-1}n^{1/(d+1)+\delta_3} \to 0$, $N|\triangle_\eta|^2 \to \infty$, $n^2|\triangle_\eta|^4/\log n \to \infty$.

The above assumptions are mild conditions that can be satisfied in many practical situations. Assumption (A1) is typically assumed about the true underlying functions in the nonparametric estimation literature. Assumption (A1) can be relaxed by only requiring $\mu(\cdot) \in \mathcal{C}^{(0)}(\Omega)$ if the imaging data has sharp edges. Assumptions (A2) and (A3) are common conditions used in the literature; see for example, Cao et al. (2012). Assumption (A4) suggests the use of more uniform triangulations with smaller shape parameters. Assumption (A5) describes the requirement of the growth rate of the dimension of the spline spaces relative to the sample size and the image resolution.

B.2 Properties of Bivariate Splines

For $g_1(\boldsymbol{z})$, $g_1(\boldsymbol{z})$, define the theoretical and empirical inner products as

$$\langle g_1, g_2 \rangle = \int_{\Omega} g_1(\boldsymbol{z}) g_2(\boldsymbol{z}) d\boldsymbol{z}, \ \langle g_1, g_2 \rangle_N = \frac{1}{N} \sum_{j=1}^{N} g_1(\boldsymbol{z}_j) g_2(\boldsymbol{z}_j), \tag{B.1}$$

and denote the corresponding theoretical and empirical norms $\|\cdot\|$ and $\|\cdot\|_N$. Furthermore, let $\|\cdot\|_{\mathcal{E}}$ be the norm introduced by the inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}}$, where, for $g_1(\boldsymbol{z})$ and $g_2(\boldsymbol{z})$,

$$\langle g_1, g_2 \rangle_{\mathcal{E}} = \int_{\Omega} \left\{ \sum_{i+j=2} \binom{2}{i} (\nabla_{z_1}^{i} \nabla_{z_2}^{j} g_1(\boldsymbol{z}))^2 \right\}^{1/2} \left\{ \sum_{i+j=2} \binom{2}{i} (\nabla_{z_1}^{i} \nabla_{z_2}^{j} g_2(\boldsymbol{z}))^2 \right\}^{1/2} dz_1 dz_2.$$

Let $A(\Omega)$ be the area of the domain Ω, and without loss of generality, we assume $A(\Omega) = 1$ in the rest of the article.

We cite two important results from Lai and Schumaker (2007).

Lemma B.1 (Theorem 2.7, Lai and Schumaker (2007)) *Let $\{B_m\}_{m \in \mathcal{M}}$ be the Bernstein polynomial basis for spline space $\mathcal{S}_d^r(\triangle)$ defined over a π-quasi-uniform triangulation $\triangle$. Then there exist positive constants c, C depending on the smoothness r, d, and the shape parameter π such that*

$$c|\triangle|^2 \sum_{m \in \mathcal{M}} \gamma_m^2 \leq \left\| \sum_{m \in \mathcal{M}} \gamma_m B_m \right\|^2 \leq C|\triangle|^2 \sum_{m \in \mathcal{M}} \gamma_m^2.$$

Lemma B.2 (Theorems 10.2 and 10.10, Lai and Schumaker (2007)) *Suppose that $\triangle$ is a π-quasi-uniform triangulation of a polygonal domian Ω, and $g(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$.*

(i) For bi-integer (a_1, a_2) with $0 \leq a_1 + a_2 \leq d$, there exists a spline $g^*(\cdot) \in \mathcal{S}_d^0(\triangle)$ such that $\|\nabla_{z_1}^{a_1} \nabla_{z_2}^{a_2} (g - g^*)\|_\infty \leq C|\triangle|^{d+1-a_1-a_2} |g|_{d+1,\infty}$, where C is a constant depending on d, and the shape parameter π.

(ii) For bi-integer (a_1, a_2) with $0 \leq a_1 + a_2 \leq d$, there exists a spline $g^{**}(\cdot) \in \mathcal{S}_d^r(\triangle)$ $(d \geq 3r+2)$ such that $\|\nabla_{z_1}^{a_1} \nabla_{z_2}^{a_2} (g - g^{**})\|_\infty \leq C|\triangle|^{d+1-a_1-a_2}|g|_{d+1,\infty}$, where C is a constant depending on d, r, and the shape parameter π.

Lemma B.2 shows that $\mathcal{S}_d^0(\triangle)$ has full approximation power, and $\mathcal{S}_d^r(\triangle)$ also has full approximation power if $d \geq 3r + 2$.

Lemma B.3 (Lemma B.4 in Supplemental Materials, Yu et al. (2019)) *Under Assumptions (A3) and (A4), for any Bernstein basis polynomials $B_m(\boldsymbol{z}),\ m \in \mathcal{M}$, of degree $d \geq 0$, one has*

$$\max_{m \in \mathcal{M}} \left| \frac{1}{N} \sum_{j=1}^{N} B_m^k(\boldsymbol{z}_j) - \int_\Omega B_m^k(\boldsymbol{z}) d\boldsymbol{z} \right| = O(N^{-1/2}|\triangle|),\ \ 1 \leq k < \infty,$$

$$\max_{m,m' \in \mathcal{M}} \left| \frac{1}{N} \sum_{j=1}^{N} B_m(\boldsymbol{z}_j) B_{m'}(\boldsymbol{z}_j) - \int_\Omega B_m(\boldsymbol{z}) B_{m'}(\boldsymbol{z}) d\boldsymbol{z} \right| = O(N^{-1/2}|\triangle|),\ \ 1 \leq k < \infty,$$

$$\max_{m,m' \in \mathcal{M}} \left| \frac{1}{N^2} \sum_{j,j'=1}^{N} G_\eta(\boldsymbol{z}_j, \boldsymbol{z}_{j'}) B_m(\boldsymbol{z}_j) B_{m'}(\boldsymbol{z}_{j'}) - \int_{\Omega^2} G_\eta(\boldsymbol{z}, \boldsymbol{z}') B_m(\boldsymbol{z}) B_{m'}(\boldsymbol{z}') d\boldsymbol{z} d\boldsymbol{z}' \right| = O(N^{-1/2}|\triangle|^3),$$

$$\max_{m \in \mathcal{M}} \left| \|\sigma B_m\|_N^2 - \|\sigma B_m\|^2 \right| = \max_{m \in \mathcal{M}} \left| \frac{1}{N} \sum_{j=1}^{N} B_m^2(\boldsymbol{z}_j) \sigma^2(\boldsymbol{z}_j) - \int_\Omega \sigma^2(\boldsymbol{z}) B_m^2(\boldsymbol{z}) d\boldsymbol{z} \right| = O(N^{-1/2}|\triangle|).$$

The following lemma provides the uniform convergence rate at which the empirical inner product approximates the theoretical inner product defined in (B.1).

Lemma B.4 *Let $g_1(\boldsymbol{z}) = \sum_{m \in \mathcal{M}} \gamma_{1,m} B_m(\boldsymbol{z})$, $g_2(\boldsymbol{z}) = \sum_{m \in \mathcal{M}} \gamma_{2,m} B_m(\boldsymbol{z})$ be any spline functions in $\mathcal{S}_d^r(\triangle)$. Suppose Assumptions (A1), (A2) and (A4) hold, and $N^{1/2}|\triangle| \to \infty$ as $N \to \infty$, then*

$$\omega_N = \sup_{g_1, g_2 \in \mathcal{S}_d^r(\triangle)} \left| \frac{\langle g_1, g_2 \rangle_N - \langle g_1, g_2 \rangle}{\|g_1\| \, \|g_2\|} \right| = O_P\left(N^{-1/2}|\triangle|^{-1}\right) = o_P(1).$$

Proof. It is easy to see

$$\langle g_1, g_2\rangle_N = \frac{1}{N}\sum_{j=1}^{N}\left\{\sum_{m\in\mathcal{M}}\gamma_{1,m}B_m(\boldsymbol{z}_j)\right\}\left\{\sum_{m'\in\mathcal{M}}\gamma_{2,m'}B_{m'}(\boldsymbol{z}_j)\right\}$$
$$= \sum_{m}\sum_{m'}\gamma_{1,m}\gamma_{2,m'}\frac{1}{N}\sum_{j=1}^{N}B_m(\boldsymbol{z}_j)B_{m'}(\boldsymbol{z}_j).$$

Note that $\langle g_1, g_2\rangle = \sum_m\sum_{m'}\gamma_{1,m}\gamma_{2,m'}\int_\Omega B_m(\boldsymbol{z})B_{m'}(\boldsymbol{z})d\boldsymbol{z}$. It follows from Assumptions (A1), (A2) and Lemma B.1 that, for any $l = 1, 2$, $\widetilde{c}_l|\triangle|^2\sum_m\gamma_{l,m}^2 \le \|g_l\|^2 \le \widetilde{C}_l|\triangle|^2\sum_m\gamma_{l,m}^2$, and

$$C_1|\triangle|^2\left(\sum_m\gamma_{1,m}^2\sum_{m'}\gamma_{2,m'}^2\right)^{1/2} \le \|g_1\|\|g_2\| \le C_2|\triangle|^2\left(\sum_m\gamma_{1,m}^2\sum_{m'}\gamma_{2,m'}^2\right)^{1/2}.$$

Therefore, one has

$$\omega_N \le \frac{\sum_{|m'-m|\le(d+2)(d+1)/2}|\gamma_{1,m}\gamma_{2,m'}|}{C_1|\triangle|^2\left[\sum_m\gamma_{1,m}^2\sum_{m'}\gamma_{2,m'}^2\right]^{1/2}}\max_{m,m'\in\mathcal{M}}\left|\frac{1}{N}\sum_{j=1}^{N}B_m(\boldsymbol{z}_j)B_{m'}(\boldsymbol{z}_j) - \int_\Omega B_m(\boldsymbol{z})B_{m'}(\boldsymbol{z})d\boldsymbol{z}\right|$$
$$\le C|\triangle|^{-2}\max_{m,m'\in\mathcal{M}}\left|\frac{1}{N}\sum_{j=1}^{N}B_m(\boldsymbol{z}_j)B_{m'}(\boldsymbol{z}_j) - \int_\Omega B_m(\boldsymbol{z})B_{m'}(\boldsymbol{z})d\boldsymbol{z}\right|.$$

The desired result follows from Lemma B.3. ■

As a direct result of Lemma B.4, we can see that

$$\sup_{g\in\mathcal{S}_d^r(\triangle)}\left|\,\|g\|_N^2/\,\|g\|^2 - 1\right| = O_P\left(N^{-1/2}|\triangle|^{-1}\right) = o_P(1). \tag{B.2}$$

Lemma B.5 *Suppose Assumption (A4) hold, and $N^{1/2}|\triangle| \to \infty$ as $N \to \infty$, then*

$$S_N = \sup_{g\in\mathcal{S}_d^r(\triangle)}\left\{\frac{\|g\|_\infty}{\|g\|_N}, \|g\|_N \ne 0\right\} = O(|\triangle|^{-1}), \tag{B.3}$$

$$\overline{S}_N = \sup_{g\in\mathcal{S}_d^r(\triangle)}\left\{\frac{\|g\|_\mathcal{E}}{\|g\|_N}, \|g\|_N \ne 0\right\} = O(|\triangle|^{-2}). \tag{B.4}$$

Proof. By Markov's inequality, for any $g \in \mathcal{S}_d^r(\triangle)$, $\|g\|_\infty \le C|\triangle|^{-1}\|g\|$, $\|g\|_\mathcal{E} \le C|\triangle|^{-2}\|g\|$.

Equation (B.2) implies that $\|g\|_N/\|g\| \geq \left[1 - O_P\left\{N^{-1/2}|\triangle|^{-1}\right\}\right]^{1/2}$. Thus, one has

$$S_N \leq C|\triangle|^{-1}\left[1 - O_P\left\{N^{-1/2}|\triangle|^{-1}\right\}\right]^{-1/2} = O_P\left(|\triangle|^{-1}\right),$$

$$\overline{S}_N \leq C|\triangle|^{-2}\left[1 - O_P\left\{N^{-1/2}|\triangle|^{-1}\right\}\right]^{-1/2} = O_P\left(|\triangle|^{-2}\right).$$

Lemma B.5 is established. ■

B.3 Convergence of Penalized Spline Estimators

Let $\left\{\widetilde{B}_m(\boldsymbol{z}), m \in \widetilde{\mathcal{M}}\right\}$ be a set of transformed Bernstein basis polynomials and $\widetilde{\mathbf{B}}(\boldsymbol{z}) = \mathbf{Q}_2^\top \mathbf{B}(\boldsymbol{z})$, then, for $\mathbf{U} = \mathbf{B}\mathbf{Q}_2$ defined in Section 2.1, $\mathbf{U}^\top\mathbf{U} = \sum_{j=1}^N \widetilde{\mathbf{B}}(\boldsymbol{z}_j)\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)$ and $\mathbf{U}^\top\mathbf{Y} = \sum_{j=1}^N \widetilde{\mathbf{B}}(\boldsymbol{z}_j)Y_{ij}$.

Denote by

$$\boldsymbol{\Gamma}_{N,\rho} = \frac{1}{N}\sum_{j=1}^N \{\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)\} + \frac{\rho_n}{nN}\mathbf{Q}_2^\top[\langle B_m, B_{m'}\rangle_{\mathcal{E}}]_{m,m'\in\mathcal{M}}\mathbf{Q}_2 \tag{B.5}$$

a symmetric positive definite matrix.

The following lemma shows that the maximum and minimum eigenvalue of $\boldsymbol{\Gamma}_{N,\rho}$ are bounded by certain orders.

Lemma B.6 *Under Assumption (A4), if $N^{1/2}|\triangle| \to \infty$ as $N \to \infty$, then there exist constants $0 < c_\rho < C_\rho < \infty$, such that with probability approaching 1 as $N \to \infty$, $n \to \infty$,*

$$c_\rho|\triangle|^2 \leq \lambda_{\min}(\boldsymbol{\Gamma}_{N,\rho}) \leq \lambda_{\max}(\boldsymbol{\Gamma}_{N,\rho}) \leq C_\rho\left(|\triangle|^2 + \frac{\rho_n}{nN|\triangle|^2}\right).$$

Specifically, when $\rho_n = 0$, one has $c_0|\triangle|^2 \leq \lambda_{\min}(\boldsymbol{\Gamma}_{N,0}) \leq \lambda_{\max}(\boldsymbol{\Gamma}_{N,0}) \leq C_0|\triangle|^2$.

Proof. For any vector $\boldsymbol{\theta}$ with the same dimension as that of $\widetilde{\mathbf{B}}(\boldsymbol{z})$, there exists $h \in \mathcal{S}_d^r(\triangle)$ such that $h(\boldsymbol{z}) = \widetilde{\mathbf{B}}^\top(\boldsymbol{z})\boldsymbol{\theta} = \mathbf{B}^\top(\boldsymbol{z})\boldsymbol{\gamma}$, where $\boldsymbol{\gamma} = \mathbf{Q}_2\boldsymbol{\theta}$ and

$$\boldsymbol{\theta}^\top\boldsymbol{\Gamma}_{N,\rho}\boldsymbol{\theta} = \frac{1}{N}\boldsymbol{\gamma}^\top\sum_{j=1}^N\{\mathbf{B}(\boldsymbol{z}_j)\mathbf{B}^\top(\boldsymbol{z}_j)\}\boldsymbol{\gamma} + \frac{\rho_n}{nN}\boldsymbol{\gamma}^\top[\langle B_m, B_{m'}\rangle_{\mathcal{E}}]_{m,m'\in\mathcal{M}}\boldsymbol{\gamma} = \|h\|_N^2 + \frac{\rho_n}{nN}\|h\|_{\mathcal{E}}^2.$$

By (B.2) and Lemma B.1, one has $\left|\|h\|_N^2/\|h\|^2-1\right| \leq \omega_N$ and

$$c(1-\omega_N)|\triangle|^2\|\boldsymbol{\gamma}\|^2 \leq (1-\omega_N)\|h\|^2 \leq \|h\|_N^2 = (1+\omega_N)\|h\|^2 \leq C(1+\omega_N)|\triangle|^2\|\boldsymbol{\gamma}\|^2.$$

Thus, $\lambda_{\min}(\boldsymbol{\Gamma}_{N,\rho}) \geq c_\rho|\triangle|^2$ for some positive constant c_ρ.

On the other hand, similar as in the supplement of Lai and Wang (2013), using the Markov's inequality and Lemma B.1, one has $\|h\|_{\mathcal{E}}^2 \leq C|\triangle|^{-4}\|h\|^2 \leq C|\triangle|^{-2}\|\boldsymbol{\gamma}\|^2$. Thus, the largest eigenvalue of the matrix $\boldsymbol{\Gamma}_{N,\rho}$ in (B.5) satisfies that

$$\lambda_{\max}(\boldsymbol{\Gamma}_{N,\rho}) \leq C\left\{(1+\omega_N)|\triangle|^2 + \frac{\rho_n}{nN}\frac{1}{|\triangle|^2}\right\} \leq C_\rho\left(|\triangle|^2 + \frac{\rho_n}{nN|\triangle|^2}\right),$$

for some positive constant C_ρ. ■

Using $\boldsymbol{\Gamma}_{N,\rho}$ defined in (B.5), the solution of the penalized regression problem (3) is given by

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\Gamma}_{N,\rho}^{-1}\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)Y_{ij}.$$

Next we define

$$\widehat{\boldsymbol{\theta}}_\mu = \boldsymbol{\Gamma}_{N,\rho}^{-1}\frac{1}{N}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\mu(\boldsymbol{z}_j),\ \ \widehat{\boldsymbol{\theta}}_\eta = \boldsymbol{\Gamma}_{N,\rho}^{-1}\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\sum_{k=1}^{\infty}\xi_{ik}\phi_k(\boldsymbol{z}_j),$$
$$\widehat{\boldsymbol{\theta}}_\varepsilon = \boldsymbol{\Gamma}_{N,\rho}^{-1}\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij}. \tag{B.6}$$

Note that, the BPS estimator $\widehat{\mu}$ in Section 2.1 can be written as $\widehat{\mu}(\boldsymbol{z}) = \widehat{\mu}^o(\boldsymbol{z}) + \widehat{\eta}(\boldsymbol{z}) + \widehat{\varepsilon}(\boldsymbol{z})$, where

$$\widehat{\mu}^o(\boldsymbol{z}) = \widetilde{\mathbf{B}}(\boldsymbol{z})^\top\widehat{\boldsymbol{\theta}}_\mu,\ \ \widehat{\eta}(\boldsymbol{z}) = \widetilde{\mathbf{B}}(\boldsymbol{z})^\top\widehat{\boldsymbol{\theta}}_\eta,\ \ \widehat{\varepsilon}(\boldsymbol{z}) = \widetilde{\mathbf{B}}(\boldsymbol{z})^\top\widehat{\boldsymbol{\theta}}_\varepsilon, \tag{B.7}$$

Therefore,

$$\widehat{\mu}(\boldsymbol{z}) - \mu(\boldsymbol{z}) = \widehat{\mu}^o(\boldsymbol{z}) - \mu(\boldsymbol{z}) + \widehat{\eta}(\boldsymbol{z}) + \widehat{\varepsilon}(\boldsymbol{z}). \tag{B.8}$$

Lemma B.7 *Suppose Assumptions (A2)–(A4) hold and* $N^{1/2}|\triangle| \to \infty$ *as* $N \to \infty$*, then* $\|\widehat{\boldsymbol{\theta}}_\eta\|^2 = O_P(n^{-1}|\triangle|^{-2})$ *and* $\|\widehat{\boldsymbol{\theta}}_\varepsilon\|^2 = O_P(n^{-1}N^{-1}|\triangle|^{-4})$.

Proof. Note that

$$\widehat{\boldsymbol{\theta}}_{\eta}=\boldsymbol{\Gamma}_{n,\rho}^{-1}\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\sum_{k=1}^{\infty}\xi_{ik}\phi_k(\boldsymbol{z}_j).$$

By Lemma B.6, one has

$$\|\widehat{\boldsymbol{\theta}}_{\eta}\|^2\asymp\frac{1}{n^2N^2|\triangle|^4}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)^{\top}\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\sum_{k,k'=1}^{\infty}\xi_{ik}\phi_k(\boldsymbol{z}_j)\xi_{i'k'}\phi_{k'}(\boldsymbol{z}_{j'}).$$

Note that by Assumption (A2), for any $i\neq i'$, j, j', one has

$$E\Big\{\widetilde{\mathbf{B}}(\boldsymbol{z}_j)^{\top}\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\sum_{k,k'=1}^{\infty}\xi_{ik}\phi_k(\boldsymbol{z}_j)\xi_{i'k'}\phi_{k'}(\boldsymbol{z}_{j'})\Big\}=\sum_{m\in\widetilde{\mathcal{M}}}\widetilde{B}_m(\boldsymbol{z}_j)\widetilde{B}_m(\boldsymbol{z}_{j'})\sum_{k,k'}E\xi_{ik}\xi_{i'k'}\phi_k(\boldsymbol{z}_j)\phi_{k'}(\boldsymbol{z}_{j'})=0.$$

Next, for any i, because $\widetilde{\mathbf{B}}(\boldsymbol{z}_j)^{\top}\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})=\mathbf{B}(\boldsymbol{z}_j)^{\top}\mathbf{Q}_2\mathbf{Q}_2^{\top}\mathbf{B}(\boldsymbol{z}_{j'})$ and the eigenvalues of $\mathbf{Q}_2\mathbf{Q}_2^{\top}$ are either 0 or 1,

$$\begin{aligned}&\frac{1}{N^2}\sum_{j=1}^{N}\sum_{j'=1}^{N}E\left\{\widetilde{\mathbf{B}}(\boldsymbol{z}_j)^{\top}\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\sum_{k,k'=1}^{\infty}\xi_{ik}\phi_k(\boldsymbol{z}_j)\xi_{ik'}\phi_{k'}(\boldsymbol{z}_{j'})\right\}\\&\quad\leq\frac{1}{N^2}\sum_{j=1}^{N}\sum_{j'=1}^{N}E\left\{\mathbf{B}(\boldsymbol{z}_j)^{\top}\mathbf{B}(\boldsymbol{z}_{j'})\sum_{k=1}^{\infty}\xi_{ik}^2\phi_k(\boldsymbol{z}_j)\phi_k(\boldsymbol{z}_{j'})\right\}\\&\quad=\sum_{m\in\mathcal{M}}\frac{1}{N^2}\sum_{j=1}^{N}\sum_{j'=1}^{N}B_m(\boldsymbol{z}_j)B_m(\boldsymbol{z}_{j'})G_{\eta}(\boldsymbol{z}_j,\boldsymbol{z}_{j'}).\end{aligned}$$

Assumption (A4) and Lemma B.3 imply that

$$\begin{aligned}\frac{1}{N^2}\sum_{j\neq j'}B_m(\boldsymbol{z}_j)B_m(\boldsymbol{z}_{j'})G_{\eta}(\boldsymbol{z}_j,\boldsymbol{z}_{j'})&=\int_{\Omega^2}G_{\eta}(\boldsymbol{z},\boldsymbol{z}')B_m(\boldsymbol{z})B_m(\boldsymbol{z}')d\boldsymbol{z}d\boldsymbol{z}'\\&\qquad\times\{1+O(N^{-1/2}|\triangle|^3)\}=O(|\triangle|^4).\end{aligned}$$

Thus,

$$\frac{1}{N^2}\sum_{j=1}^{N}\sum_{j'=1}^{N}E\left\{\widetilde{\mathbf{B}}(\boldsymbol{z}_j)^{\top}\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\sum_{k,k'=1}^{\infty}\xi_{ik}\phi_k(\boldsymbol{z}_j)\xi_{ik'}\phi_{k'}(\boldsymbol{z}_{j'})\right\}\leq C|\triangle|^2.$$

Therefore, $E\|\widehat{\boldsymbol{\theta}}_{\eta}\|^2\leq C(n^{-1}|\triangle|^{-2})$.

Similarly, by the definition of $\widehat{\boldsymbol{\theta}}_\varepsilon$ in (B.6) and Lemma B.6, one has

$$\|\widehat{\boldsymbol{\theta}}_\varepsilon\|^2 \asymp \frac{1}{n^2N^2|\triangle|^4}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)^\top\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\sigma(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_{j'})\varepsilon_{ij}\varepsilon_{i'j}.$$

Note that for any $i \neq i'$, j, j', $E(\varepsilon_{ij}\varepsilon_{i'j'}) = 0$ and for any $i, j \neq j'$, $E(\varepsilon_{ij}\varepsilon_{ij'}) = 0$. Because the eigenvalues of $\mathbf{Q}_2\mathbf{Q}_2^\top$ are either 0 or 1, by Assumption (A2) and Lemma B.3, for any i,

$$E\Big\{\frac{1}{N}\sum_{j,j'=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)^\top\widetilde{\mathbf{B}}(\boldsymbol{z}_{j'})\sigma(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_{j'})\varepsilon_{ij}\varepsilon_{i'j}\Big\} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{B}(\boldsymbol{z}_j)^\top\mathbf{Q}_2\mathbf{Q}_2^\top\mathbf{B}(\boldsymbol{z}_j)\sigma^2(\boldsymbol{z}_j)$$
$$\le C\sum_{m\in\mathcal{M}}\frac{1}{N}\sum_{j=1}^{N}B_m^2(\boldsymbol{z}_j)\sigma^2(\boldsymbol{z}_j) \le C\sum_{m\in\mathcal{M}}\int_\Omega \sigma^2(\boldsymbol{z})B_m^2(\boldsymbol{z})d\boldsymbol{z}\{1+O(N^{-1/2}|\triangle|^{-1})\} = O(1).$$

Therefore,

$$E\|\widehat{\boldsymbol{\theta}}_\varepsilon\|^2 \asymp \frac{1}{nN|\triangle|^4}\frac{1}{N}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)^\top\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\sigma^2(\boldsymbol{z}_j) \le C(nN)^{-1}|\triangle|^{-4}.$$

The conclusion of the lemma follows. ■

Next, the following lemmas give the uniform convergence rate of $\widehat{\mu}(\boldsymbol{z})$ to $\mu(\boldsymbol{z})$. We start by introducing some notations for the specific situation when there is no penalty in the regression problem, i.e., $\rho_n = 0$. Denote $\boldsymbol{\Gamma}_{N,0} = \frac{1}{N}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\widetilde{\mathbf{B}}^\top(\boldsymbol{z}_j)$. Let $\overline{\xi}_{\cdot k} = \frac{1}{n}\sum_{i=1}^{n}\xi_{ik}$, for any $k \ge 1$, and $\overline{\varepsilon}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ij}$ for any $j = 1, \ldots, N$, and denote

$$\widetilde{\boldsymbol{\theta}}_\mu = \boldsymbol{\Gamma}_{N,0}^{-1}\frac{1}{N}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\mu(\boldsymbol{z}_j),$$
$$\widetilde{\boldsymbol{\theta}}_\eta = \boldsymbol{\Gamma}_{N,0}^{-1}\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\eta_i(\boldsymbol{z}_j) = \boldsymbol{\Gamma}_{N,0}^{-1}\frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{\kappa}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\overline{\xi}_{\cdot k}\phi_k(\boldsymbol{z_j}),$$
$$\widetilde{\boldsymbol{\theta}}_\varepsilon = \boldsymbol{\Gamma}_{N,0}^{-1}\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij} = \boldsymbol{\Gamma}_{N,0}^{-1}\frac{1}{N}\sum_{j=1}^{N}\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\overline{\varepsilon}_{\cdot j},$$

Then we can have the following decomposition $\widetilde{\mu}(\boldsymbol{z}) = \widetilde{\mu}^o(\boldsymbol{z}) + \widetilde{\eta}(\boldsymbol{z}) + \widetilde{\varepsilon}(\boldsymbol{z})$, where

$$\widetilde{\mu}^o(\boldsymbol{z}) = \widetilde{\mathbf{B}}(\boldsymbol{z})^\top\widetilde{\boldsymbol{\theta}}_\mu,\ \widetilde{\eta}(\boldsymbol{z}) = \widetilde{\mathbf{B}}(\boldsymbol{z})^\top\widetilde{\boldsymbol{\theta}}_\eta,\ \widetilde{\varepsilon}(\boldsymbol{z}) = \widetilde{\mathbf{B}}(\boldsymbol{z})^\top\widetilde{\boldsymbol{\theta}}_\varepsilon. \tag{B.9}$$

Lemma B.8 *Under Assumptions (A1) and (A4), if* $N^{1/2}|\triangle| \to \infty$ *as* $N \to \infty$*, the functions* $\widehat{\mu}^o(\boldsymbol{z})$ *satisfy* $\|\widehat{\mu}^o - \mu\|_\infty = O_P\left\{\frac{\rho_n}{nN|\triangle|^3}|\mu|_{2,\infty} + \left(1 + \frac{\rho_n}{nN|\triangle|^5}\right)|\triangle|^{d+1}|\mu|_{d+1,\infty}\right\}$.

Proof. Note that $\|\mu - \widehat{\mu}^o\|_\infty \leq \|\mu - \widetilde{\mu}^o\|_\infty + \|\widetilde{\mu}^o - \widehat{\mu}^o\|_\infty$, where $\widetilde{\mu}^o$ is given in (B.9). According to Proposition 1 in Lai and Wang (2013), $\|\widetilde{\mu}^o - \mu\|_\infty \leq C|\triangle|^{d+1}|\mu|_{d+1,\infty}$. Thus we only need to show the order of $\|\widetilde{\mu}^o - \widehat{\mu}^o\|_\infty$.

By the definition of S_N in (B.3), one has

$$\|\widetilde{\mu}^o - \widehat{\mu}^o\|_\infty \leq S_N\|\widetilde{\mu}^o - \widehat{\mu}^o\|_N. \tag{B.10}$$

Note that the penalized spline $\widehat{\mu}^o$ of μ is characterized by the orthogonality relation: $nN\langle\mu - \widehat{\mu}^o, g\rangle_N = \rho_n\langle\widehat{\mu}^o, g\rangle_{\mathcal{E}}, \quad$ for all $g \in \mathcal{S}_d^r(\triangle)$, while $\widetilde{\mu}^o$ is characterized by $\langle\mu - \widetilde{\mu}^o, g\rangle_N = 0$, for all $g \in \mathcal{S}_d^r(\triangle)$. Combining the two orthogonality relations, one has $nN\langle\widetilde{\mu}^o - \widehat{\mu}^o, g\rangle_N = \rho_n\langle\widehat{\mu}^o, g\rangle_{\mathcal{E}}$, for all $g \in \mathcal{S}_d^r(\triangle)$. Inserting $g = \widetilde{\mu}^o - \widehat{\mu}^o$ yields that

$$nN\|\widetilde{\mu}^o - \widehat{\mu}^o\|_N^2 = \rho_n\langle\widehat{\mu}^o, \widetilde{\mu}^o - \widehat{\mu}^o\rangle_{\mathcal{E}} = \rho_n\left\{\langle\widehat{\mu}^o, \widetilde{\mu}^o\rangle_{\mathcal{E}} - \|\widehat{\mu}^o\|_{\mathcal{E}}^2\right\} \geq 0.$$

Thus, by Cauchy-Schwarz inequality, $\|\widehat{\mu}^o\|_{\mathcal{E}}^2 \leq \langle\widehat{\mu}^o, \widetilde{\mu}^o\rangle_{\mathcal{E}} \leq \|\widehat{\mu}^o\|_{\mathcal{E}}\|\widetilde{\mu}^o\|_{\mathcal{E}}$, which implies that $\|\widehat{\mu}^o\|_{\mathcal{E}} \leq \|\widetilde{\mu}^o\|_{\mathcal{E}}$. Meanwhile, by the definition of $\overline{S}_N$,

$$nN\|\widetilde{\mu}^o - \widehat{\mu}^o\|_N^2 \leq \rho_n\|\widehat{\mu}^o\|_{\mathcal{E}}\|\widetilde{\mu}^o - \widehat{\mu}^o\|_{\mathcal{E}} \leq \rho_n\overline{S}_N\|\widehat{\mu}^o\|_{\mathcal{E}}\|\widetilde{\mu}^o - \widehat{\mu}^o\|_N \leq \rho_n\overline{S}_N\|\widetilde{\mu}^o\|_{\mathcal{E}}\|\widetilde{\mu}^o - \widehat{\mu}^o\|_N.$$

Therefore,

$$\|\widetilde{\mu}^o - \widehat{\mu}^o\|_N \leq \rho_n(nN)^{-1}\overline{S}_N\|\widetilde{\mu}^o\|_{\mathcal{E}}. \tag{B.11}$$

Combining (B.10) and (B.11) yields that $\|\widetilde{\mu}^o - \widehat{\mu}^o\|_\infty \leq S_N\|\widetilde{\mu}^o - \widehat{\mu}^o\|_N \leq \rho_n(nN)^{-1}S_N\overline{S}_N\|\widetilde{\mu}^o\|_{\mathcal{E}}$. By Lemma B.2, one has

$$\|\widetilde{\mu}^o\|_{\mathcal{E}} = C_1\{|\mu|_{2,\infty} + \sum_{a_1+a_2=2}\|\nabla_{z_1}^{a_1}\nabla_{z_2}^{a_2}(\mu - \widetilde{\mu}^o)\|_\infty\} \leq C_2(|\mu|_{2,\infty} + |\triangle|^{d-1}|\mu|_{d+1,\infty}).$$

It follows $\|\widetilde{\mu}^o - \widehat{\mu}^o\|_\infty = \rho_n (nN)^{-1} S_N \overline{S}_N C_2 (|\mu|_{2,\infty} + |\triangle|^{d-1}|\mu|_{d+1,\infty})$. By Lemma B.5, one has $S_N = O_P(|\triangle|^{-1})$ and $\overline{S}_N = O_P(|\triangle|^{-2})$. Thus,

$$\|\widetilde{\mu}^o - \widehat{\mu}^o\|_\infty = O_P\left\{\frac{\rho_n}{nN|\triangle|^3}(|\mu|_{2,\infty} + |\triangle|^{d-1}|\mu|_{d+1,\infty})\right\}.$$

Hence, $\|\widehat{\mu}^o - \mu\|_\infty \le C_1|\triangle|^{d+1}|\mu|_{d+1,\infty} + O_P\left\{\frac{\rho_n}{nN|\triangle|^3}\left(|\mu|_{2,\infty} + |\triangle|^{d-1}|\mu|_{d+1,\infty}\right)\right\}$. Lemma B.8 is established. ■

Lemma B.9 *Under Assumptions (A2)–(A4), if* $N^{1/2}|\triangle| \to \infty$ *as* $N \to \infty$ *and* $n^{1/(4+\delta_2)} \ll n^{1/2}N^{-1/2}|\triangle|^{-1}$*, then* $\|\widetilde{\varepsilon}\|_\infty = O_P\{(nN)^{-1/2}(\log n)^{1/2}|\triangle|^{-1}\}$*. In addition, if Assumption (A6) holds, then* $\|\widetilde{\eta}\|_\infty = O_P\{n^{-1/2}(\log n)^{1/2}\}$.

Proof. Note that $\widetilde{\varepsilon}(\boldsymbol{z}) = \sum_{m\in\widetilde{\mathcal{M}}} \widetilde{\theta}_{\varepsilon,m}\widetilde{B}_m(\boldsymbol{z})$ for some coefficients $\widetilde{\theta}_{\varepsilon,m}$, so the order of $\widehat{\varepsilon}(\boldsymbol{z})$ is related to that of $\widetilde{\theta}_{\varepsilon,m}$. In fact

$$\begin{aligned}\|\widetilde{\varepsilon}\|_\infty &= \left\|\widetilde{\mathbf{B}}(\boldsymbol{z})^\top \boldsymbol{\Gamma}_{N,0}^{-1}\left[\frac{1}{nN}\sum_{i=1}^n\sum_{j=1}^N \widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij}\right]_{m\in\widetilde{\mathcal{M}}}\right\|_\infty \\ &\le C|\triangle|^{-2}\max_{m\in\widetilde{\mathcal{M}}}\left|\frac{1}{nN}\sum_{i=1}^n\sum_{j=1}^N \widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij}\right|,\end{aligned}$$

almost surely, where $\widetilde{\boldsymbol{\theta}}_\varepsilon = (\widetilde{\theta}_{\varepsilon,m})_{m\in\widetilde{\mathcal{M}}}$ with $\widetilde{\mathcal{M}}$ being an index set of the transformed Bernstein basis polynomials $\widetilde{B}_m(\boldsymbol{z})$. Next we show that with probability 1

$$\max_{m\in\widetilde{\mathcal{M}}}\left|\frac{1}{nN}\sum_{i=1}^n\sum_{j=1}^N \widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij}\right| = O\left\{(\log n)^{1/2}|\triangle|/(nN)^{1/2}\right\}. \quad \text{(B.12)}$$

To prove (B.12), let $\tau_i = \tau_{i,m} = \frac{1}{nN}\sum_{j=1}^N \widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij}$. We decompose the random variable τ_i into a truncated part and a tail part,

$$\tau_{i,1}^{L_n} = \frac{1}{nN}\sum_{j=1}^N \widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij} I\{|\varepsilon_{ij}| > L_n\}, \ \tau_{i,2}^{L_n} = \frac{1}{nN}\sum_{j=1}^N \widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij} I\{|\varepsilon_{ij}| \le L_n\} - \mu_i^{L_n},$$

$$\mu_i^{L_n} = \frac{1}{nN}\sum_{j=1}^N \widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)E\left[\varepsilon_{ij} I\{|\varepsilon_{ij}| \le L_n\}\right],$$

where $L_n = n^{\alpha}$, and $n^{1/(4+\delta_2)} \ll n^{\alpha} \ll \sqrt{\frac{n}{N \log n}} |\triangle|^{-1}$.

It is straightforward to verify that $\mu_i^{L_n} = O(n^{-1} L_n^{-2} |\triangle|^2)$. Next we show that tail part vanishes almost surely. Note that

$$\sum_{n=1}^{\infty} P\{|\varepsilon_{nj}| > L_n\} \le \sum_{n=1}^{\infty} \frac{E|\varepsilon_{nj}|^{4+\delta_2}}{L_n^{4+\delta_2}} \le \upsilon_{\delta} \sum_{n=1}^{\infty} L_n^{-(4+\delta_2)} < \infty. \tag{B.13}$$

By Borel Cantelli lemma, one has $\left|\sum_{i=1}^{n} \tau_{i,1}^{L_n}\right| = O_{a.s.}\left(n^{-k}\right)$, for any $k > 0$. Next, note that $E\left(\tau_{i,2}^{L_n}\right) = 0$, one has

$$\begin{aligned} \operatorname{Var}(\tau_{i,2}^{L_n}) &= \frac{1}{n^2 N^2} \sum_{j=1}^{N} \widetilde{B}_m^2(\boldsymbol{z}_j) \sigma^2(\boldsymbol{z}_j) \left\{ E\left(\varepsilon_{ij}^2\right) - E[\varepsilon_{ij}^2 I\{|\varepsilon_{ij}| > L_n\}] - \left(E\left[\varepsilon_{ij} I\left\{|\varepsilon_{ij}| \le L_n\right\}\right]\right)^2 \right\} \\ &\asymp n^{-2} N^{-1} |\triangle|^2. \end{aligned}$$

Using the independence of $\tau_{i,2}^{L_n}$, $i = 1, \ldots, n$, one has $\operatorname{Var}\left(\sum_{i=1}^{n} \tau_{i,2}^{L_n}\right) \asymp (nN)^{-1} |\triangle|^2$.

Now Minkowski's inequality implies that

$$\begin{aligned} E\left|\tau_{i,2}^{L_n}\right|^k &= E\left|\frac{1}{nN} \sum_{j=1}^{N} \widetilde{B}_m(\boldsymbol{z}_j) \sigma(\boldsymbol{z}_j) \varepsilon_{ij} I\left\{|\varepsilon_{ij}| \le L_n\right\} - \mu_i^{L_n}\right|^k \\ &\le 2^{k-1} \left[\left\{ \frac{1}{nN} \sum_{j=1}^{N} \widetilde{B}_m(\boldsymbol{z}_j) \sigma(\boldsymbol{z}_j) L_n \right\}^{k-2} E\left|\frac{1}{nN} \sum_{j=1}^{N} \widetilde{B}_m(\boldsymbol{z}_j) \sigma(\boldsymbol{z}_j) \varepsilon_{ij} I\left\{|\varepsilon_{ij}| \le L_n\right\}\right|^2 + \left(\mu_i^{L_n}\right)^k \right] \\ &\le 2 \left(\frac{C|\triangle|^2 L_n}{n}\right)^{k-2} E\left|\tau_{i,2}^{L_n}\right|^2, \quad k \ge 3. \end{aligned}$$

Thus, $E\left|\tau_{i,2}^{L_n}\right|^k \le \left(\frac{C L_n |\triangle|^2}{n}\right)^{k-2} k! E|\tau_{i,2}^{L_n}|^2 < \infty$ with the Cramer constant $c^* = C n^{-1} L_n |\triangle|^2$.

By the Bernstein inequality, for any large enough $\delta > 0$,

$$P\left(\left|\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij}\right| \geq \delta|\triangle|\sqrt{\frac{\log n}{nN}}\right) = P\left(\left|\sum_{i=1}^{n}\tau_{i,2}^{L_n}\right| \geq \delta|\triangle|\sqrt{\frac{\log n}{nN}}\right)$$
$$\leq 2\exp\left\{\frac{-\delta^2|\triangle|^2\frac{\log n}{nN}}{4\mathrm{Var}\left(\sum_{i=1}^{n}\tau_{i,2}^{L_n}\right) + 2c^*\delta|\triangle|\sqrt{\frac{\log n}{nN}}}\right\} = 2\exp\left\{\frac{-\delta^2|\triangle|^2\frac{\log n}{nN}}{\frac{4c}{nN}|\triangle|^2 + 2CL_n n^{-1}\delta|\triangle|^3\sqrt{\frac{\log n}{nN}}}\right\}$$
$$= 2\exp\left\{\frac{-\delta^2\log n}{4c + 2CL_n\delta|\triangle|\sqrt{\frac{N\log n}{n}}}\right\} \leq 2n^{-3},$$

given that $L_n = n^{\alpha} = o\left(\sqrt{\frac{n}{N\log n}}|\triangle|^{-1}\right)$. Hence

$$\sum_{n=1}^{\infty} P\left(\max_{m\in\mathcal{M}}\left|\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\widetilde{B}_m(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij}\right| \geq \delta|\triangle|\sqrt{\frac{\log n}{nN}}\right) \leq c|\triangle|^{-2}\sum_{n=1}^{\infty} n^{-3} < \infty,$$

for such $\delta > 0$. Thus, Borel-Cantelli's lemma implies (B.12).

Similarly, for $\widetilde{\eta}(\boldsymbol{z}) = \sum_{m\in\widetilde{\mathcal{M}}}\widetilde{\theta}_{\eta,m}\widetilde{B}_m(\boldsymbol{z})$ one has

$$\|\widetilde{\eta}\|_{\infty} \leq C|\triangle|^{-2}\max_{m\in\widetilde{\mathcal{M}}}\left|\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\eta_i(\boldsymbol{z}_j)\widetilde{B}_m(\boldsymbol{z}_j)\right|,$$

almost surely. Then we can show that with probability 1

$$\max_{m\in\widetilde{\mathcal{M}}}\left|\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\eta_i(\boldsymbol{z}_j)\widetilde{B}_m(\boldsymbol{z}_j)\right| = O\left\{n^{-1/2}|\triangle|^2(\log n)^{1/2}\right\},$$

by decomposing mean 0 random variable $u_i = u_{i,m} = \frac{1}{N}\sum_{j=1}^{N}\eta_i(\boldsymbol{z}_j)\widetilde{B}_m(\boldsymbol{z}_j)$ into

$$u_{i,1}^{T_n} = \sum_{k=1}^{\infty}\left\{\frac{1}{nN}\sum_{j=1}^{N}\widetilde{B}_m(\boldsymbol{z}_j)\phi_k(\boldsymbol{z}_j)\right\}\xi_{ik}I\left\{|\xi_{ik}| > T_n\right\},$$
$$u_{i,2}^{T_n} = \sum_{k=1}^{\infty}\left\{\frac{1}{nN}\sum_{j=1}^{N}\widetilde{B}_m(\boldsymbol{z}_j)\phi_k(\boldsymbol{z}_j)\right\}\xi_{ik}I\left\{|\xi_{ik}| \leq T_n\right\} - \mu_i^{T_n},$$
$$\mu_i^{T_n} = \sum_{k=1}^{\infty}\left\{\frac{1}{nN}\sum_{j=1}^{N}\widetilde{B}_m(\boldsymbol{z}_j)\phi_k(\boldsymbol{z}_j)\right\}E\left[\xi_{ik}I\left\{|\xi_{ik}| \leq T_n\right\}\right],$$

where $T_n = n^{\alpha}$ and $n^{1/(4+\delta_1)} \ll n^{\alpha} \ll (n/\log n)^{1/2}$.

Using Borel Cantelli lemma and similar method in (B.13), we can show that tail part vanishes almost surely, i.e., $\left|\sum_{i=1}^{n} u_{i,1}^{T_n}\right| = O_{a.s.}\left(n^{-r}\right),$ for any $r > 0$. As $Eu_i = 0$, then it is straightforward to verify that $\mu_i^{T_n} = -Eu_{i,1}^{T_n} = O(n^{-1}T_n^{-2}|\triangle|^2)$.

Next, notice that $Eu_{i,2}^{T_n} = 0$. Then, one has

$$\begin{aligned}\mathrm{Var}\left(u_{i,2}^{T_n}\right) =& \frac{1}{n^2N^2}\sum_{k=1}^{\infty}\left\{\sum_{j=1}^{N}\sum_{j'=1}^{N}\widetilde{B}_m(\boldsymbol{z}_j)\widetilde{B}_m(\boldsymbol{z}_{j'})\phi_k(\boldsymbol{z}_l)\phi_k(\boldsymbol{z}_{j'})\right\}\\ &\times\left\{E(\xi_{ik}^2) - E\Big[\xi_{ik}^2 I\{|\xi_{ik}| > T_n\}\Big] - \left(E\left[\xi_{ik} I\left\{|\xi_{ik}| \leq T_n\right\}\right]\right)^2\right\} = O(n^{-2}|\triangle|^4),\end{aligned}$$

which indicates $\mathrm{Var}\left(\sum_{i=1}^{n} u_{i,2}^{T_n}\right) = n^{-1}|\triangle|^4$.

Similarly, we can show that there exists some constant C, such that for any $r \geq 3$, we have $E|u_{i,2}^{T_n}|^r \leq \left(C|\triangle|^2T_n/n\right)^{r-2} r!E|u_{i,2}^{T_n}|^2$. Using Bernstein inequality, one has

$$P\left\{\left|\sum_{i=1}^{n} u_i\right| \geq \delta n^{-1/2}|\triangle|^2(\log n)^{1/2}\right\} \leq 2\exp\left\{\frac{-\delta^2\log n}{4c + 2\delta C T_n(\log n)^{1/2}n^{-1/2}}\right\} \leq 2n^{-3}.$$

Hence,

$$\sum_{n=1}^{\infty} P\left\{\max_{m\in\widetilde{\mathcal{M}}}\left|\sum_{i=1}^{n} u_i\right| \geq \delta n^{-1/2}|\triangle|^2(\log n)^{1/2}\right\} \leq C|\triangle|^{-2}\sum_{n=1}^{\infty} n^{-3} < \infty$$

for such $\delta > 0$. Thus, Borel-Cantelli's lemma implies that $\|\widetilde{\eta}\|_\infty = O_P\{n^{-1/2}(\log n)^{1/2}\}$. ■

Lemma B.10 *Under Assumptions (A2)–(A4), one has*

$$\|\widehat{\varepsilon}\|_\infty = O_P\left\{\frac{(\log n)^{1/2}}{\sqrt{nN}|\triangle|} + \frac{\rho_n}{n^{3/2}N^{3/2}|\triangle|^6}\right\},\ \|\widehat{\eta}\|_\infty = O_P\left\{\frac{(\log n)^{1/2}}{\sqrt{n}} + \frac{\rho_n}{n^{3/2}N|\triangle|^5}\right\}.$$

Proof. We only show the infinity norm of $\widehat{\varepsilon}$. The conclusion of $\|\widehat{\eta}\|_\infty$ follows similarly. Note that the penalized spline $\widehat{\varepsilon}$ of ε is characterized by the orthogonality relations: $nN\left\langle \varepsilon - \widehat{\varepsilon}, g\right\rangle_N = \rho_n\left\langle \widehat{\varepsilon}, g\right\rangle_{\mathcal{E}}$, for all $g \in \mathcal{S}_d^r(\triangle)$. In particular, $\widetilde{\varepsilon}$ is characterized by $\left\langle \varepsilon - \widetilde{\varepsilon}, g\right\rangle_N = 0$, for all $g \in \mathcal{S}_d^r(\triangle)$. Inserting $g = \widetilde{\varepsilon} - \widehat{\varepsilon}$ yield that $nN\left\|\widetilde{\varepsilon} - \widehat{\varepsilon}\right\|_N^2 = \rho_n\left\langle \widehat{\varepsilon}, \widetilde{\varepsilon} - \widehat{\varepsilon}\right\rangle_{\mathcal{E}} = \rho_n(\langle\widehat{\varepsilon}, \widetilde{\varepsilon}\rangle_{\mathcal{E}} - \|\widehat{\varepsilon}\|_{\mathcal{E}})$.

It follows, by Cauchy-Schwarz inequality, that $\|\widehat{\varepsilon}\|_{\mathcal{E}}^2 \leq \langle \widehat{\varepsilon}, \widetilde{\varepsilon} \rangle_{\mathcal{E}} \leq \|\widehat{\varepsilon}\|_{\mathcal{E}} \|\widetilde{\varepsilon}\|_{\mathcal{E}}$, which implies that $\|\widehat{\varepsilon}\|_{\mathcal{E}} \leq \|\widetilde{\varepsilon}\|_{\mathcal{E}}$. Thus, by Cauchy-Schwarz inequality and the definition of $\overline{S}_N$ in (B.4), one has

$$nN \|\widetilde{\varepsilon} - \widehat{\varepsilon}\|_N^2 \leq \rho_n \|\widehat{\varepsilon}\|_{\mathcal{E}} \|\widetilde{\varepsilon} - \widehat{\varepsilon}\|_{\mathcal{E}} \leq \overline{S}_N \rho_n \|\widehat{\varepsilon}\|_{\mathcal{E}} \|\widetilde{\varepsilon} - \widehat{\varepsilon}\|_N .$$

Hence, $\|\widetilde{\varepsilon} - \widehat{\varepsilon}\|_N \leq (nN)^{-1} \overline{S}_N \rho_n \|\widetilde{\varepsilon}\|_{\mathcal{E}}$. Using (B.3), we obtain

$$\|\widetilde{\varepsilon} - \widehat{\varepsilon}\|_\infty \leq S_N \|\widetilde{\varepsilon} - \widehat{\varepsilon}\|_N \leq (nN)^{-1} S_N \overline{S}_N \rho_n \|\widetilde{\varepsilon}\|_{\mathcal{E}} .$$

Finally, we use Markov's inequality to get $\|\widetilde{\varepsilon}\|_{\mathcal{E}} \leq C_1 |\triangle|^{-2} \|\widetilde{\varepsilon}\|$. It therefore follows that

$$\|\widehat{\varepsilon}\|_\infty \leq \|\widetilde{\varepsilon}\|_\infty + \|\widetilde{\varepsilon} - \widehat{\varepsilon}\|_\infty \leq \|\widetilde{\varepsilon}\|_\infty + \frac{\rho_n}{nN} S_N \overline{S}_N \frac{C_1}{|\triangle|^2} \|\widetilde{\varepsilon}\|_{L_2}.$$

According to Lemmas B.5, B.7 and B.9, one has $\|\widetilde{\varepsilon}\|_\infty = O_P\left\{(nN)^{-1/2} (\log n)^{1/2} |\triangle|^{-1}\right\}$ and $\|\widehat{\varepsilon}\|_{L_2}^2 \asymp |\triangle|^2 \|\widehat{\boldsymbol{\theta}}_\varepsilon\|^2 = O_P(n^{-1} N^{-1} |\triangle|^{-2})$. Hence, $\|\widehat{\varepsilon}\|_\infty = O_P\left\{\frac{(\log n)^{1/2}}{\sqrt{nN}|\triangle|} + \frac{\rho_n}{n^{3/2} N^{3/2} |\triangle|^6}\right\}$. ■

Proof of Theorem 1. Note that $\widehat{\mu} - \mu = \widehat{\mu}^o - \mu + \widehat{\eta} + \widehat{\varepsilon}$. Therefore, $\|\widehat{\mu} - \mu\|_{L_2} \leq \|\widehat{\mu}^o - \mu\|_{L_2} + \|\widehat{\eta}\|_{L_2} + \|\widehat{\varepsilon}\|_{L_2}$. By Lemmas B.1 and B.7, one has

$$\|\widehat{\eta}\|_{L_2}^2 \asymp |\triangle|^2 \|\widehat{\boldsymbol{\theta}}_\eta\|^2 = O_P(n^{-1}), \ \|\widehat{\varepsilon}\|_{L_2}^2 \asymp |\triangle|^2 \|\widehat{\boldsymbol{\theta}}_\varepsilon\|^2 = O_P(n^{-1} N^{-1} |\triangle|^{-2}),$$

and the asymptotic order of $\|\widehat{\mu}^o - \mu\|_{L_2}$ is the same as $\|\widehat{\mu}^o - \mu\|_\infty$. By Lemmas B.8 and B.10,

$$\|\widehat{\mu}^o - \mu\|_\infty = O_P\left\{\frac{\rho_n}{nN|\triangle|^3} |\mu|_{2,\infty} + \left(1 + \frac{\rho_n}{nN|\triangle|^5}\right) |\triangle|^{d+1} |\mu|_{d+1,\infty}\right\},$$
$$\|\widehat{\eta}\|_\infty = O_P\left\{\frac{(\log n)^{1/2}}{\sqrt{n}} + \frac{\rho_n}{n^{3/2} N |\triangle|^5}\right\}, \ \|\widehat{\varepsilon}\|_\infty = O_P\left\{\frac{(\log n)^{1/2}}{\sqrt{nN}|\triangle|} + \frac{\rho_n}{n^{3/2} N^{3/2} |\triangle|^6}\right\}.$$

Thus, by Assumption (A5), $\|\widehat{\mu} - \mu\|_\infty = o_P\{(n^{-1} \log(n))^{1/2}\}$ and $\|\widehat{\mu} - \mu\|_{L_2} = O_P(n^{-1/2})$. ■

B.4 Simultaneous Confidence Bands

B.4.1 Proof of Theorem 3

Lemma B.11 (Lemma A.5, Cao et al. (2012)) *Let* $\overline{\xi}_{\cdot k} = \frac{1}{n} \sum_{i=1}^n \xi_{ik}$ *and* $\overline{\varepsilon}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n \varepsilon_{ij}$. *If Assumption (A2) holds, then there exists some constant* $C_\beta > 0$ *such that* $\max_{1 \leq k \leq \kappa} E|\overline{\xi}_{\cdot k} - \overline{Z}_{\cdot k,\xi}| \leq$

$C_\beta n^{\beta-1}$, and $\max_{1\le j\le N}|\bar\varepsilon_{\cdot j}-\overline{Z}_{\cdot j,\varepsilon}| = O_{a.s.}(n^{\beta-1})$, for some $\beta\in(0,1/2)$, *where* $\{Z_{ik,\xi}\}_{i=1,k=1}^{n,\kappa}$ *and* $\{Z_{ij,\varepsilon}\}_{i=1,j=1}^{n,N}$ *are iid* $N(0,1)$ *variables and* $\overline{Z}_{\cdot k,\xi} = \frac{1}{n}\sum_{i=1}^n Z_{ik,\xi}$, $\overline{Z}_{\cdot j,\varepsilon} = \frac{1}{n}\sum_{i=1}^n Z_{ij,\varepsilon}$, $1\le j\le N$, $1\le k\le\kappa$.

Lemma B.12 *Let* $\overline{\eta}(\boldsymbol{z}) = \frac{1}{n}\sum_{i=1}^n \eta_i(\boldsymbol{z}) = \sum_{k=1}^\kappa \bar\xi_{\cdot k}\phi_k(\boldsymbol{z})$, *then under Assumptions (A2)–(A6), for* $\widehat{\eta}(\boldsymbol{z})$ *defined in (B.7), one has* $n^{1/2}\|\widehat\eta-\overline\eta\|_\infty = o_P(1)$. *In addition, as* $N\to\infty$, $n\to\infty$,

$$P\left\{\sup_{\boldsymbol{z}\in\Omega} n^{1/2}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}|\overline\eta(\boldsymbol{z})|\le q_{1-\alpha}\right\}\to 1-\alpha,$$
$$P\left\{n^{1/2}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}|\overline\eta(\boldsymbol{z})|\le z_{1-\alpha/2}\right\}\to 1-\alpha,\quad \text{for any } \boldsymbol{z}\in\Omega.$$

Proof. Denote $\widetilde\zeta_k(\boldsymbol{z}) = \overline{Z}_{\cdot k,\xi}\phi_k(\boldsymbol{z})$, $k=1,\ldots,\kappa$, and define

$$\widetilde\zeta(\boldsymbol{z}) = n^{1/2}\left\{\sum_{k=1}^\kappa \phi_k^2(\boldsymbol{z})\right\}^{-1/2}\sum_{k=1}^\kappa\widetilde\zeta_k(\boldsymbol{z}) = n^{1/2}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}\sum_{k=1}^\kappa\widetilde\zeta_k(\boldsymbol{z}),$$

then $\{\widetilde\zeta(\boldsymbol{z}),\boldsymbol{z}\in\Omega\}$ is a Gaussian random field with mean 0, variance 1 and covariance function $\mathrm{Cov}\left\{\widetilde\zeta(\boldsymbol{z}),\widetilde\zeta(\boldsymbol{z}')\right\} = G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}G_\eta(\boldsymbol{z},\boldsymbol{z}')G_\eta(\boldsymbol{z}',\boldsymbol{z}')^{-1/2}$. Therefore, $\widetilde\zeta(\boldsymbol{z})$ has the same distribution as $\zeta(\boldsymbol{z}),\boldsymbol{z}\in\Omega$.

Next, let $\widehat\phi_k(\boldsymbol{z}) = \widetilde{\mathbf{B}}(\boldsymbol{z})^\top\boldsymbol{\Gamma}_{N,\rho}^{-1}\frac{1}{N}\sum_{j=1}^N\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\phi_k(\boldsymbol{z}_j)$. Similar to the proof for Lemma B.9, by Lemma B.6, $\|\widehat\phi_k\|_\infty\le C|\triangle|^{-2}\left\|\frac{1}{N}\sum_{j=1}^N\widetilde{\mathbf{B}}(\boldsymbol{z}_j)\phi_k(\boldsymbol{z}_j)\right\|_\infty\le C_1\|\phi_k\|_\infty$. According to Lemma B.8,

$$\|\widehat\phi_k-\phi_k\|_\infty = O_P\left\{\frac{\rho_n}{nN|\triangle|^3}|\phi_k|_{2,\infty}+\left(1+\frac{\rho_n}{nN|\triangle|^5}\right)|\triangle|^{s+1}|\phi_k|_{s+1,\infty}\right\}.$$

Therefore, by Assumptions (A4)–(A6), one has

$$E\left\{n^{1/2}\sup_{\boldsymbol{z}\in\Omega}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}|\widehat\eta(\boldsymbol{z})-\overline\eta(\boldsymbol{z})|\right\} = E\left[n^{1/2}\sup_{\boldsymbol{z}\in\Omega}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}\left|\sum_{k=1}^\kappa\bar\xi_{\cdot k}\{\phi_k(\boldsymbol{z})-\widehat\phi_k(\boldsymbol{z})\}\right|\right]$$
$$\le n^{1/2}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}\left(\sum_{k=1}^{\kappa_n}\|\widehat\phi_k-\phi_k\|_\infty E|\bar\xi_{\cdot k}|+C\sum_{k=\kappa_n+1}^{\kappa}\|\phi_k\|_\infty E|\bar\xi_{\cdot k}|\right).$$

Thus,

$$
\begin{aligned}
&E\Big\{n^{1/2}\sup_{\boldsymbol{z}\in\Omega}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}|\widehat{\eta}(\boldsymbol{z})-\overline{\eta}(\boldsymbol{z})|\Big\}\\
\leq& n^{1/2}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}C_1\left[\sum_{k=1}^{\kappa_n}\left\{\frac{\rho_n}{nN|\triangle|^3}|\phi_k|_{2,\infty}+\left(1+\frac{\rho_n}{nN|\triangle|^5}\right)|\triangle|^{s+1}|\phi_k|_{s+1,\infty}\right\}\right.\\
&\left.\times E|\overline{\xi}_{\cdot k}|+\sum_{k=\kappa_n+1}^{\kappa}E|\overline{\xi}_{\cdot k}|\|\phi_k\|_\infty\right]\\
\leq& C_2\left\{\frac{\rho_n}{nN|\triangle|^3}\sum_{k=1}^{\kappa_n}|\phi_k|_{2,\infty}+\left(1+\frac{\rho_n}{nN|\triangle|^5}\right)\sum_{k=1}^{\kappa_n}|\triangle|^{s+1}|\phi_k|_{s+1,\infty}+\sum_{k=\kappa_n+1}^{\kappa}\|\phi_k\|_\infty\right\}=o(1).
\end{aligned}
$$

Hence, $n^{1/2}\sup_{\boldsymbol{z}\in\Omega}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}|\widehat{\eta}(\boldsymbol{z})-\overline{\eta}(\boldsymbol{z})|=o_P(1)$. Under Assumption (A3), it follows that $\|\widehat{\eta}-\overline{\eta}\|_\infty=o_P(n^{-1/2})$.

By Lemma B.11, for some $\beta\in(0,1/2)$,

$$
\begin{aligned}
E\Big\{\sup_{\boldsymbol{z}\in\Omega}\Big|\widetilde{\zeta}(\boldsymbol{z})-n^{1/2}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}\overline{\eta}(\boldsymbol{z})\Big|\Big\}&=E\left\{n^{1/2}\sup_{\boldsymbol{z}\in\Omega}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}\left|\sum_{k=1}^{\kappa}(\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k})\phi_k(\boldsymbol{z})\right|\right\}\\
&\leq Cn^{1/2}\sum_{k=1}^{\kappa}\|\phi_k\|_\infty E|\overline{Z}_{\cdot k,\xi}-\overline{\xi}_{\cdot k}|\leq\widetilde{C}n^{\beta-1/2}\sum_{k=1}^{\kappa}\|\phi_k\|_\infty.
\end{aligned}
$$

Thus, by Assumption (A6), $\sup_{\boldsymbol{z}\in\Omega}\left|\widetilde{\zeta}(\boldsymbol{z})-n^{1/2}G_\eta(\boldsymbol{z},\boldsymbol{z})^{-1/2}\overline{\eta}(\boldsymbol{z})\right|=o_P(1)$. Finally, note that $P\left\{\sup_{\boldsymbol{z}\in\Omega}\left|\widetilde{\zeta}(\boldsymbol{z})\right|\leq q_{1-\alpha}\right\}=1-\alpha$. The lemma is proved. ■

Proof of Theorem 3. Note that "oracle" estimator $\overline{\mu}(\boldsymbol{z})=\mu(\boldsymbol{z})+\overline{\eta}(\boldsymbol{z})$ implies that $\widehat{\mu}-\overline{\mu}=\widehat{\mu}^o-\mu+\widehat{\eta}-\overline{\eta}+\widehat{\varepsilon}$. By Lemmas B.8, B.10, Assumptions (A5) and (A6),

$$
\begin{aligned}
\|\widehat{\mu}^o-\mu\|_\infty&=O_P\left\{\frac{\rho_n}{nN|\triangle|^3}|\mu|_{2,\infty}+\left(1+\frac{\rho_n}{nN|\triangle|^5}\right)|\triangle|^{d+1}|\mu|_{d+1,\infty}\right\}=o_P(n^{-1/2}),\\
\|\widehat{\varepsilon}\|_\infty&=O_P\left\{\frac{(\log n)^{1/2}}{\sqrt{nN}|\triangle|}+\frac{\rho_n}{n^{3/2}N^{3/2}|\triangle|^6}\right\}=o_P(n^{-1/2}).
\end{aligned}
$$

Thus, according to Lemma B.12, the theorem is established. ■

B.4.2 Proof of Theorem 4

According to (B.8), for $H = 1, 2$, we can decompose the unpenalized spline estimator $\widehat{\mu}_H(\cdot)$ as $\widehat{\mu}_H(\boldsymbol{z}) = \widehat{\mu}_H^o(\boldsymbol{z}) + \widehat{\eta}_H(\boldsymbol{z}) + \widehat{\varepsilon}_H(\boldsymbol{z})$. Therefore, asymptotic error $(\widehat{\mu}_1 - \widehat{\mu}_2) - (\mu_1 - \mu_2)$ can be decomposed into three components: $(\widehat{\mu}_1^o - \widehat{\mu}_2^o - \mu_1 + \mu_2) + (\widehat{\eta}_1 - \widehat{\eta}_2) + (\widehat{\varepsilon}_1 - \widehat{\varepsilon}_2)$. Similar as the proof of Theorem 2, the first and third components of the decomposition can be proved to have $\sqrt{n}$ asymptotic efficiency. Here we focus on the second component.

By Lemma B.11, one can find i.i.d $Z_{Hik,\xi} \sim N(0,1)$, $i = 1, \ldots, n_H$ such that $\max_{1 \le k \le \kappa_H} E|\overline{\xi}_{H\cdot k} - \overline{Z}_{H\cdot k,\xi}| \le C_0 n^{\beta-1}$ and $\overline{Z}_{H\cdot k,\xi} = n_H^{-1} \sum_{i=1}^{n_H} Z_{Hik,\xi}$. Likewise, for the white noise sequence $\{\varepsilon_{Hij}, i \ge 1\}$, one can also find iid $Z_{Hik,\varepsilon} \sim N(0,1)$, $i = 1, \ldots, n_H$ such that $\max_{1 \le j \le N} |\overline{\varepsilon}_{H\cdot j} - \overline{Z}_{H\cdot j,\varepsilon}| = O_{a.s.}(n^{\beta-1})$, where $\beta \in (0, 1/2)$. Let $V(\boldsymbol{z}, \boldsymbol{z}') = G_{\eta,1}(\boldsymbol{z}, \boldsymbol{z}') + \tau G_{\eta,2}(\boldsymbol{z}, \boldsymbol{z}')$, where $\tau = \lim_{n_1 \to \infty} n_1/n_2$, and define

$$\widetilde{W}(\boldsymbol{z}) = n_1^{1/2} V(\boldsymbol{z}, \boldsymbol{z})^{-1/2} \left\{ \sum_{k=1}^{\kappa_1} \overline{Z}_{1\cdot k,\xi} \phi_{1k}(\boldsymbol{z}) - \sum_{k=1}^{\kappa_2} \overline{Z}_{2\cdot k,\xi} \phi_{2k}(\boldsymbol{z}) \right\}.$$

Then, for any $\boldsymbol{z} \in \Omega$, $\widetilde{W}(\boldsymbol{z})$ is Gaussian with mean 0 and variance 1, and the covariance

$$E\left\{ \widetilde{W}(\boldsymbol{z}) \widetilde{W}(\boldsymbol{z}') \right\} = V(\boldsymbol{z}, \boldsymbol{z})^{-1/2} V(\boldsymbol{z}, \boldsymbol{z}') V(\boldsymbol{z}', \boldsymbol{z}')^{-1/2}.$$

That is, the distribution of $\widetilde{W}(\boldsymbol{z})$, $\boldsymbol{z} \in \Omega$ and the distribution of $W(\boldsymbol{z})$, $\boldsymbol{z} \in \Omega$ are identical. Similarly, for $H = 1, 2$, let $\widehat{\phi}_{Hk}(\boldsymbol{z}) = \widetilde{\mathbf{B}}_H(\boldsymbol{z})^\top \boldsymbol{\Gamma}_{H,N,\rho}^{-1} \frac{1}{N} \sum_{j=1}^N \widetilde{\mathbf{B}}_H(\boldsymbol{z}_j) \phi_{Hk}(\boldsymbol{z}_j)$. Note that

$$\overline{\eta}_H(\boldsymbol{z}) = \frac{1}{n} \sum_{i=1}^n \eta_{Hi}(\boldsymbol{z}) = \sum_{k=1}^{\kappa_H} \overline{\xi}_{H\cdot k} \phi_{Hk}(\boldsymbol{z}),\ \widehat{\eta}_H(\boldsymbol{z}) = \sum_{k=1}^{\kappa_H} \overline{\xi}_{H\cdot k} \widehat{\phi}_{Hk}(\boldsymbol{z}).$$

And we have shown in Lemma B.12 that $n_H^{1/2} \|\widehat{\eta}_H - \overline{\eta}_H\|_\infty = o_P(1)$.

Lemma B.13 *If Assumptions (A1)–(A6), are modified for each group accordingly, then one has as*

$N \to \infty$, $n_1 \to \infty$,

$$P\left\{\sup_{\boldsymbol{z}\in\Omega} n_1^{1/2} V(\boldsymbol{z},\boldsymbol{z})^{-1/2} |\overline{\eta}_1(\boldsymbol{z}) - \overline{\eta}_2(\boldsymbol{z})| \le q_{12,\alpha}\right\} \to 1-\alpha.$$

Proof. Note that, by similar discussion in Lemma B.12,

$$\begin{aligned}
&E\left[\sup_{\boldsymbol{z}\in\Omega}\left|\widetilde{W}(\boldsymbol{z}) - n_1^{1/2} V(\boldsymbol{z},\boldsymbol{z})^{-1/2}\left\{\overline{\eta}_1(\boldsymbol{z}) - \overline{\eta}_2(\boldsymbol{z})\right\}\right|\right] \\
=& n_1^{1/2} E\left[\sup_{\boldsymbol{z}\in\Omega} V(\boldsymbol{z},\boldsymbol{z})^{-1/2}\left|\sum_{k=1}^{\kappa_1}(\overline{Z}_{1\cdot k,\xi} - \overline{\xi}_{1\cdot k})\phi_{1k}(\boldsymbol{z}) - \sum_{k=1}^{\kappa_2}(\overline{Z}_{2\cdot k,\xi} - \overline{\xi}_{2\cdot k})\phi_{2k}(\boldsymbol{z})\right|\right] = o(1).
\end{aligned}$$

Therefore, one has $\sup_{\boldsymbol{z}\in\Omega}\left|\widetilde{W}(\boldsymbol{z}) - n_1^{1/2} V(\boldsymbol{z},\boldsymbol{z})^{-1/2}\left\{\overline{\eta}_1(\boldsymbol{z}) - \overline{\eta}_2(\boldsymbol{z})\right\}\right| = o_P(1)$. Observe that $P\left\{\sup_{\boldsymbol{z}\in\Omega}|\widetilde{W}(\boldsymbol{z})| \le q_{12,\alpha}\right\} = 1-\alpha$, for any $\alpha \in (0,1)$, as $N \to \infty$, $n_1 \to \infty$

$$\begin{aligned}
&P\left\{\sup_{\boldsymbol{z}\in\Omega} n_1^{1/2} V(\boldsymbol{z},\boldsymbol{z})^{-1/2} |\overline{\eta}_1(\boldsymbol{z}) - \overline{\eta}_2(\boldsymbol{z})| \le q_{12,\alpha}\right\} \to 1-\alpha, \\
&P\left\{n_1^{1/2} V(\boldsymbol{z},\boldsymbol{z})^{-1/2} |\overline{\eta}_1(\boldsymbol{z}) - \overline{\eta}_2(\boldsymbol{z})| \le z_{1-\alpha/2}\right\} \to 1-\alpha, \quad \text{for all } \boldsymbol{z} \in \Omega.
\end{aligned}$$

The conclusion of the lemma is proved. ■

B.5 Convergence of the Covariance Estimator

Without loss of generality, we prove Theorem 2 based on the unpenalized bivariate spline estimator. Using similar arguments in Section B.3, we can easily extend this proof to the penalized case. Based on the estimated residuals $\widehat{R}_{ij} = Y_{ij} - \widehat{\mu}(\boldsymbol{z}_j)$, $i = 1,\ldots,n$, $j = 1,\ldots,N$, denote $\widehat{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}} \sum_{j=1}^{N}\left\{\widehat{R}_{ij} - \mathbf{B}^*(\boldsymbol{z}_j)^\top \mathbf{Q}_2^*\boldsymbol{\beta}\right\}^2$, where $\mathbf{B}^*(\boldsymbol{z})$ is the set of bivariate spline basis functions used to estimate $\eta_i(\boldsymbol{z})$, and the transpose of $\mathbf{H}^*$ admits the following QR decomposition: $\mathbf{H}^{*\top} = \mathbf{Q}^*\mathbf{R}^* = (\mathbf{Q}_1^*\ \mathbf{Q}_2^*)\binom{\mathbf{R}_1^*}{\mathbf{R}_2^*}$. Then, the bivariate spline estimator of $\eta_i(\boldsymbol{z})$ can be written as

$$\widehat{\eta}_i(\boldsymbol{z}) = \mathbf{B}^*(\boldsymbol{z})^\top \mathbf{Q}_2^*\widehat{\boldsymbol{\beta}}_i = \widetilde{\mathbf{B}}^*(\boldsymbol{z})^\top\widehat{\boldsymbol{\beta}}_i. \tag{B.14}$$

Let $\mathbf{\Gamma^*}_N = \frac{1}{N}\sum_{j=1}^N \widetilde{\mathbf{B}}^*(\boldsymbol{z}_j)\widetilde{\mathbf{B}}^*(\boldsymbol{z}_j)^\top$, then one has

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_i &= \mathbf{\Gamma}_N^{*-1}\frac{1}{N}\sum_{j=1}^N \widetilde{\mathbf{B}}^*(\boldsymbol{z}_j)\widehat{R}_{ij} = \mathbf{\Gamma}_N^{*-1}\frac{1}{N}\sum_{j=1}^N \widetilde{\mathbf{B}}^*(\boldsymbol{z}_j)\{Y_{ij} - \widehat{\mu}(\boldsymbol{z}_j)\}\\ &= \mathbf{\Gamma}_N^{*-1}\frac{1}{N}\sum_{j=1}^N \widetilde{\mathbf{B}}^*(\boldsymbol{z}_j)\left[\{\mu(\boldsymbol{z}_j) - \widehat{\mu}(\boldsymbol{z}_j)\} + \eta_i(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij}\right].\end{aligned}$$

Next we define $\widetilde{r}(\boldsymbol{z}) = \widetilde{\mathbf{B}}^*(\boldsymbol{z})^\top\mathbf{\Gamma}_N^{*-1}\frac{1}{N}\sum_{j=1}^N \widetilde{\mathbf{B}}^*(\boldsymbol{z}_j)\{\mu(\boldsymbol{z}_j) - \widehat{\mu}(\boldsymbol{z}_j)\}$, and

$$\widetilde{\eta}_i(\boldsymbol{z}) = \widetilde{\mathbf{B}}^*(\boldsymbol{z})^\top\mathbf{\Gamma}_N^{*-1}\frac{1}{nN}\sum_{i=1}^n\sum_{j=1}^N \widetilde{\mathbf{B}}^*(\boldsymbol{z}_j)\eta_i(\boldsymbol{z}_j),\ \widetilde{\varepsilon}_i(\boldsymbol{z}) = \widetilde{\mathbf{B}}^*(\boldsymbol{z})^\top\mathbf{\Gamma}_N^{*-1}\frac{1}{nN}\sum_{i=1}^n\sum_{j=1}^N \widetilde{\mathbf{B}}^*(\boldsymbol{z}_j)\sigma(\boldsymbol{z}_j)\varepsilon_{ij}.$$

Then, the estimation error $d_i(\boldsymbol{z}) = \widehat{\eta}_i(\boldsymbol{z}) - \eta_i(\boldsymbol{z})$ in (B.14) can be decomposed as the following: $d_i(\boldsymbol{z}) = \widetilde{r}(\boldsymbol{z}) + \widetilde{\eta}_i(\boldsymbol{z}) - \eta_i(\boldsymbol{z}) + \widetilde{\varepsilon}_i(\boldsymbol{z})$.

For any $\boldsymbol{z}$, $\boldsymbol{z}' \in \Omega$, denote $\widetilde{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') = n^{-1}\sum_{i=1}^n \eta_i(\boldsymbol{z})\eta_i(\boldsymbol{z}')$. The following lemma shows the uniform convergence of $\widetilde{G}_\eta(\boldsymbol{z}, \boldsymbol{z}')$ to $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ in probability over all $(\boldsymbol{z}, \boldsymbol{z}') \in \Omega^2$.

Lemma B.14 (Lemma B.18, Yu et al. (2019)) *Under Assumptions (A1)–(A7),* $\sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2} |\widetilde{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') - G_\eta(\boldsymbol{z}, \boldsymbol{z}')| = O_P\{n^{-1/2}(\log n)^{1/2}\}$.

Proof of Theorem 2 (i). Note that

$$\sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2} |\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') - G_\eta(\boldsymbol{z}, \boldsymbol{z}')| \le \sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2}\{|\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') - \widetilde{G}_\eta(\boldsymbol{z}, \boldsymbol{z}')| + |\widetilde{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') - G_\eta(\boldsymbol{z}, \boldsymbol{z}')|\}$$

where $\sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2} |\widetilde{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') - G_\eta(\boldsymbol{z}, \boldsymbol{z}')| = o_P(1)$ according to Lemma B.14, and

$$\begin{aligned}\sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2} |\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') - \widetilde{G}_\eta(\boldsymbol{z}, \boldsymbol{z}')| &\le \sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2}\left|n^{-1}\sum_{i=1}^n \eta_i(\boldsymbol{z})d_i(\boldsymbol{z}')\right| + \sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2}\left|n^{-1}\sum_{i=1}^n \eta_i(\boldsymbol{z}')d_i(\boldsymbol{z})\right|\\ &\quad + \sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2}\left|n^{-1}\sum_{i=1}^n d_i(\boldsymbol{z})d_i(\boldsymbol{z}')\right|.\end{aligned}$$

Similar to the proof of Yu et al. (2019), one can show that

$$\sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2}\left|\frac{1}{n}\sum_{i=1}^n d_i(\boldsymbol{z})d_i(\boldsymbol{z}')\right| = o_P(1),\ \sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2}\left|\frac{1}{n}\sum_{i=1}^n \eta_i(\boldsymbol{z})d_i(\boldsymbol{z}') + n^{-1}\sum_{i=1}^n \eta_i(\boldsymbol{z}')d_i(\boldsymbol{z})\right| = o_P(1).$$

The desired result is established. ■

Proof of Theorem 2 (ii). Denote $\Delta\psi_k(\boldsymbol{z}) = \int(\widehat{G} - G)(\boldsymbol{z}, \boldsymbol{z}')\psi_k(\boldsymbol{z}')d\boldsymbol{z}'$. By Theorem 2 (i), $\|\widehat{G} - G\|_\infty = o_P(1)$. Thus, for any $k \geq 1$, $\|\Delta\psi_k\|_\infty = o_P(1)$. According to Hall and Hosseini-Nasab (2006), let $\|\mathit{\Delta}\|_2 = [\iint(\widehat{G}(\boldsymbol{z}, \boldsymbol{z}') - G(\boldsymbol{z}, \boldsymbol{z}'))^2 d\boldsymbol{z}d\boldsymbol{z}']^{1/2}$, then $\widehat{\psi}_k - \psi_k = \sum_{j\neq k}(\lambda_k - \lambda_j)^{-1}\langle\Delta\psi_k, \psi_j\rangle\psi_j + O(\|\mathit{\Delta}\|_2^2)$. It follows from Bessel's inequality that $\|\widehat{\psi}_k - \psi_k\|_2 \leq C(\|\Delta\psi_k\|_\infty^2 + O(\|\mathit{\Delta}\|_2^2)) = o_P(1)$. By (2.9) in Hall and Hosseini-Nasab (2006),

$$\widehat{\lambda}_k - \lambda_k = \iint(\widehat{G} - G)(\boldsymbol{z}, \boldsymbol{z}')\psi_k(\boldsymbol{z})\psi_k(\boldsymbol{z}')d\boldsymbol{z}d\boldsymbol{z}' + O(\|\Delta\psi_k\|_2^2).$$

Thus, using Theorem 2 (i), one has $|\widehat{\lambda}_k - \lambda_k| = o_P(1), \forall k \geq 1$.

Next, note that

$$\begin{aligned}
\widehat{\lambda}_k\widehat{\psi}_k(\boldsymbol{z}) - \lambda_k\psi_k(\boldsymbol{z}) &= \int \widehat{G}(\boldsymbol{z}, \boldsymbol{z}')\widehat{\psi}_k(\boldsymbol{z}')d\boldsymbol{z}' - \int G(\boldsymbol{z}, \boldsymbol{z}')\psi_k(\boldsymbol{z}')d\boldsymbol{z}' \\
&= \int(\widehat{G} - G)(\boldsymbol{z}, \boldsymbol{z}')(\widehat{\psi}(\boldsymbol{z}') - \psi_k(\boldsymbol{z}'))d\boldsymbol{z}' + \int(\widehat{G} - G)(\boldsymbol{z}, \boldsymbol{z}')\psi_k(\boldsymbol{z}')d\boldsymbol{z}' \\
&\quad + \int G(\boldsymbol{z}, \boldsymbol{z}')\{\widehat{\psi}_k(\boldsymbol{z}') - \psi_k(\boldsymbol{z}')\}d\boldsymbol{z}'.
\end{aligned}$$

By Cauchy-Schwarz inequality and Theorem 2 (i), for all $\boldsymbol{z} \in \Omega$,

$$\begin{aligned}
&\int G(\boldsymbol{z}, \boldsymbol{z}')\{\widehat{\psi}_k(\boldsymbol{z}') - \psi_k(\boldsymbol{z}')\}d\boldsymbol{z}' \leq \left(\int G^2(\boldsymbol{z}, \boldsymbol{z}')d\boldsymbol{z}'\right)^{1/2}\|\widehat{\psi}_k - \psi_k\|_2 = o_P(1), \\
&\int(\widehat{G} - G)(\boldsymbol{z}, \boldsymbol{z}')(\widehat{\psi}(\boldsymbol{z}') - \psi_k(\boldsymbol{z}'))d\boldsymbol{z}' \leq \|\widehat{G} - G\|_\infty\|\widehat{\psi}_k - \psi_k\|_2 = o_P(1), \\
&\int(\widehat{G} - G)(\boldsymbol{z}, \boldsymbol{z}')\psi_k(\boldsymbol{z}')d\boldsymbol{z}' \leq \|\widehat{G} - G\|_\infty\|\psi_k\|_2 = o_P(1).
\end{aligned}$$

Therefore, $\|\widehat{\lambda}_k\widehat{\psi}_k - \lambda_k\psi_k\|_\infty = o_P(1)$, and $\lambda_k\|\widehat{\psi}_k - \psi_k\|_\infty \leq \|\widehat{\lambda}_k\widehat{\psi}_k - \lambda_k\psi_k\|_\infty + |\widehat{\lambda}_k - \lambda_k|\|\widehat{\psi}_k\|_\infty = o_P(1)$. It follows that $\|\widehat{\psi}_k - \psi_k\|_\infty = o_P(1)$. ■

References

Cao, G., Yang, L., and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. Journal of nonparametric statistics **24,** 359–377.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **68,** 109–126.

Lai, M. J. and Schumaker, L. L. (2007). Spline functions on triangulations. Cambridge University Press.

Lai, M. J. and Wang, L. (2013). Bivariate penalized splines for regression. Statistica Sinica **23,** 1399–1417.

Poldrack, R. A., Mumford, J. A., and Nichols, T. E. (2011). Handbook of functional MRI data analysis. Cambridge University Press.

Yu, S., Wang, G., Wang, L., and Yang, L. (2019). Multivariate spline estimation and inference for image-on-scalar regression. Submitted.

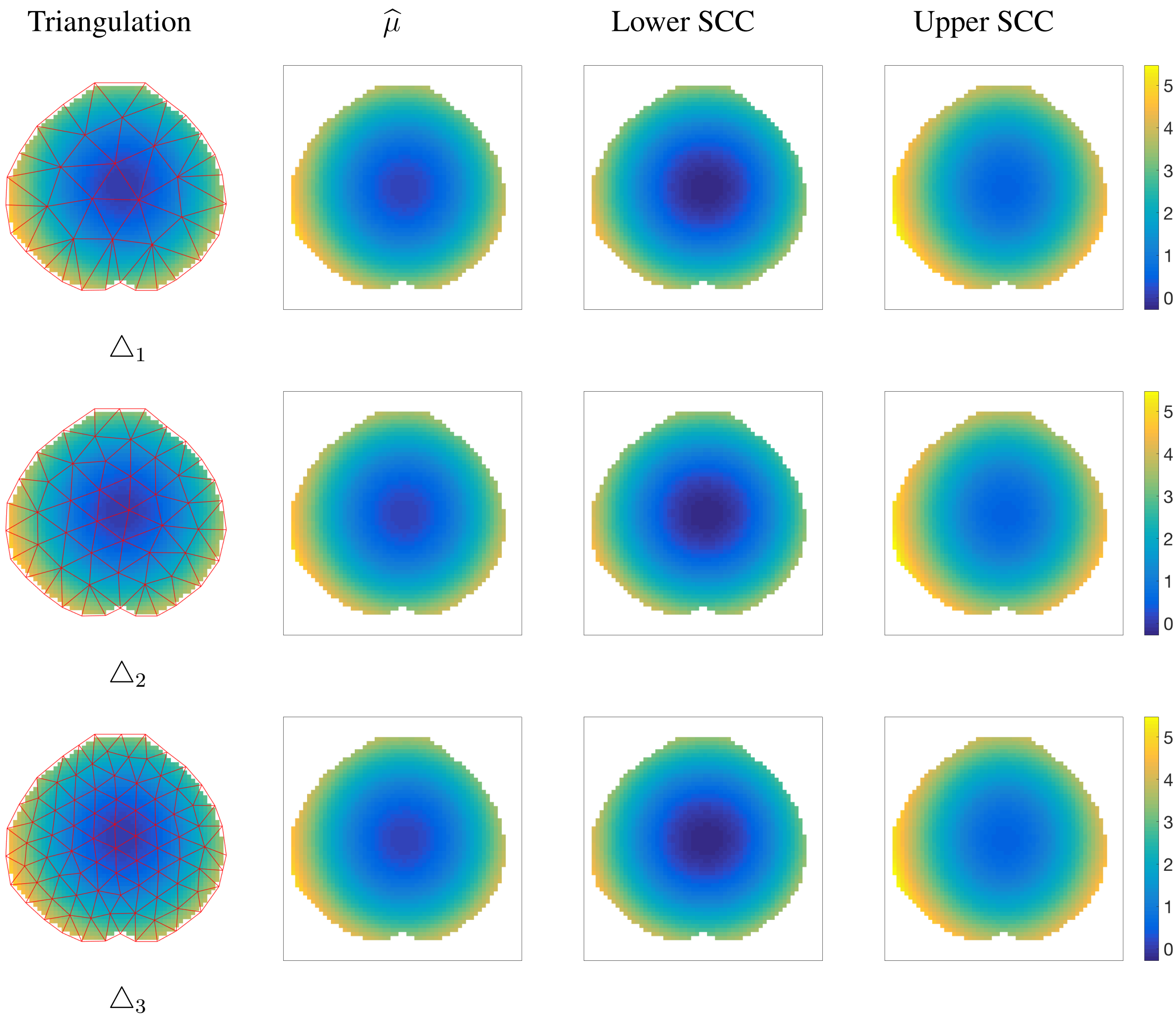


Figure A1: SCCs for quadratic function with $n = 50$ and $\alpha = 0.01$.

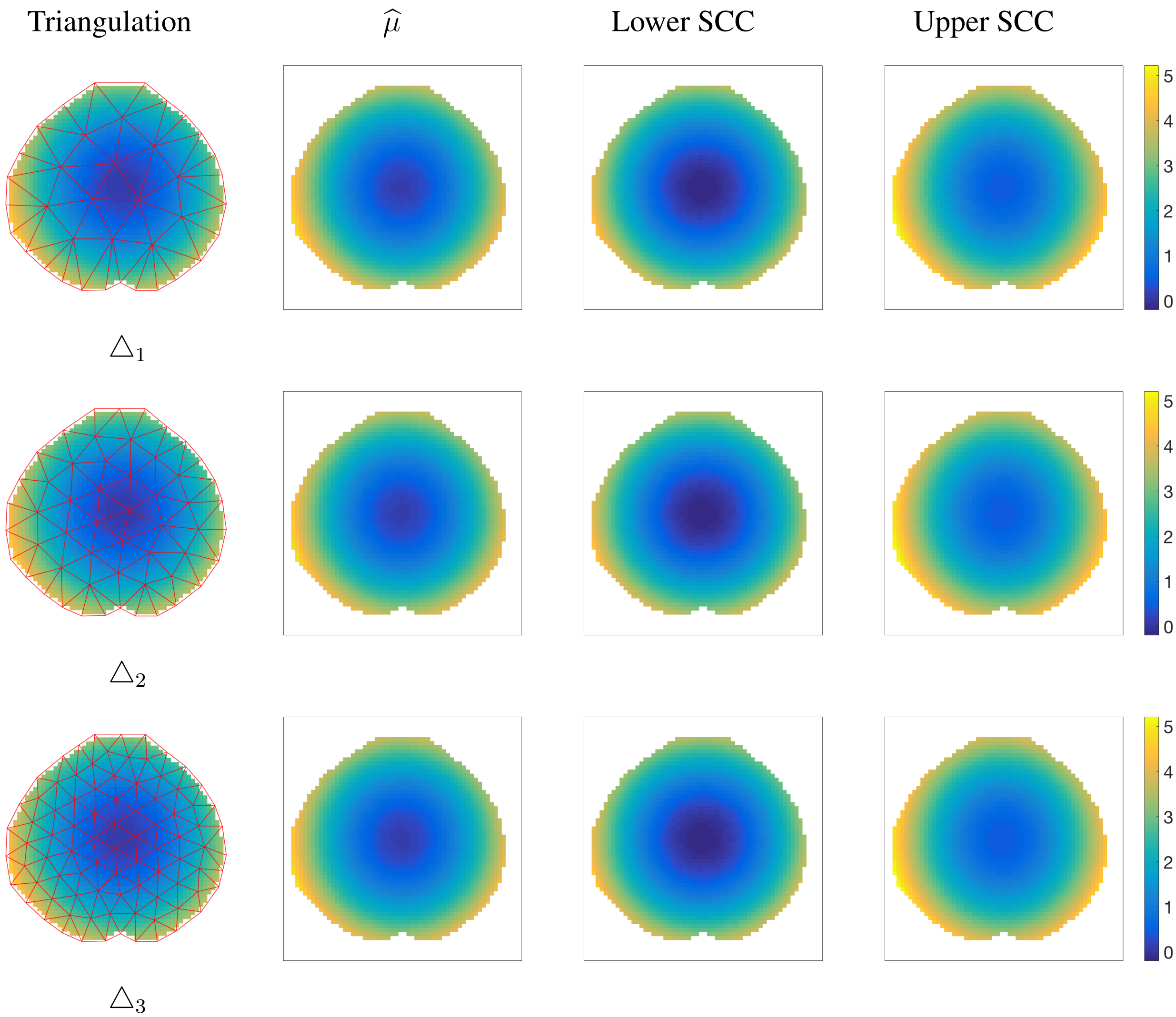


Figure A2: SCCs for quadratic function with $n = 100$ and $\alpha = 0.01$.

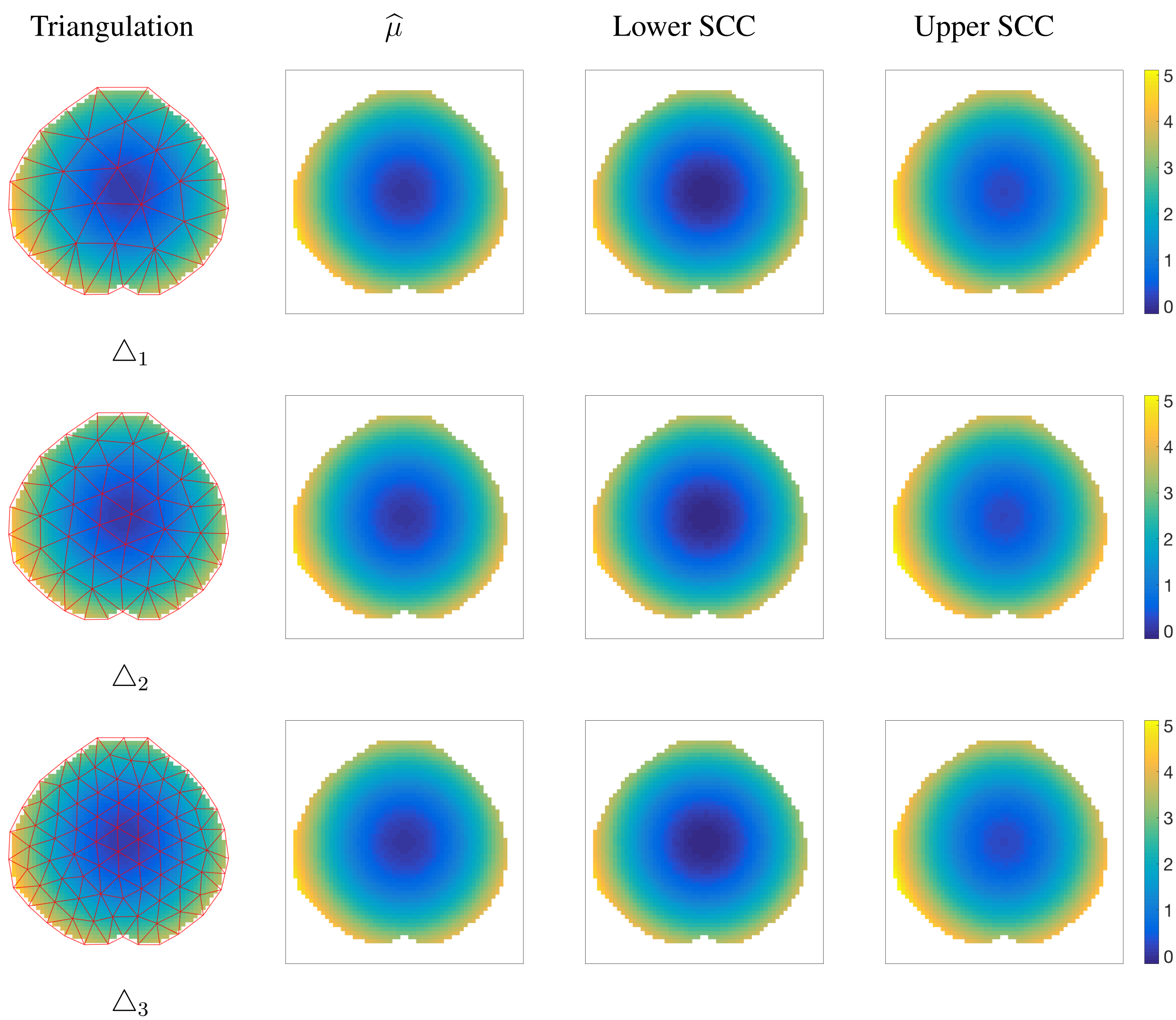


Figure A3: SCCs for quadratic function with $n = 200$ and $\alpha = 0.01$.

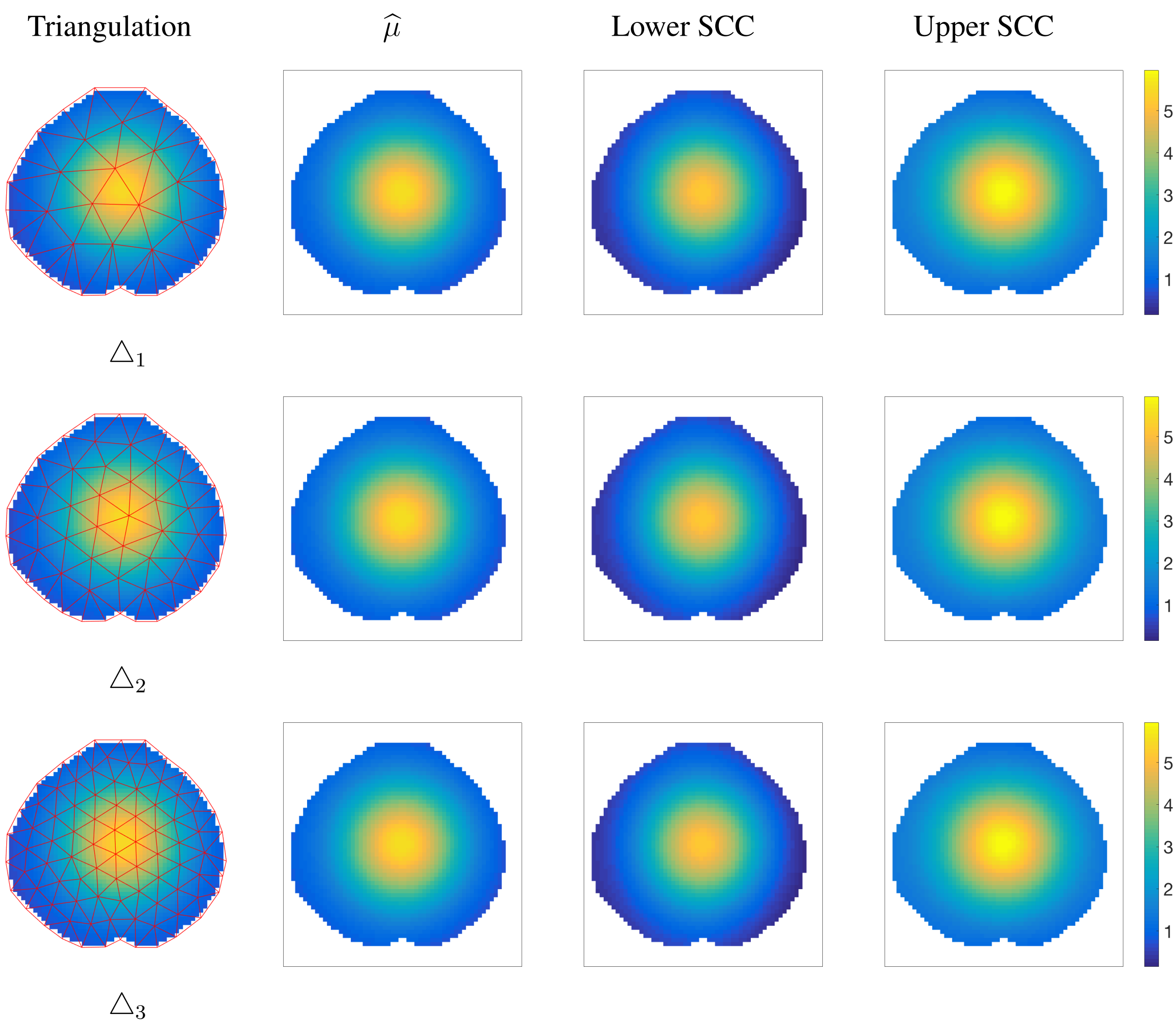


Figure A4: SCCs for bump function with $n = 50$ and $\alpha = 0.01$.

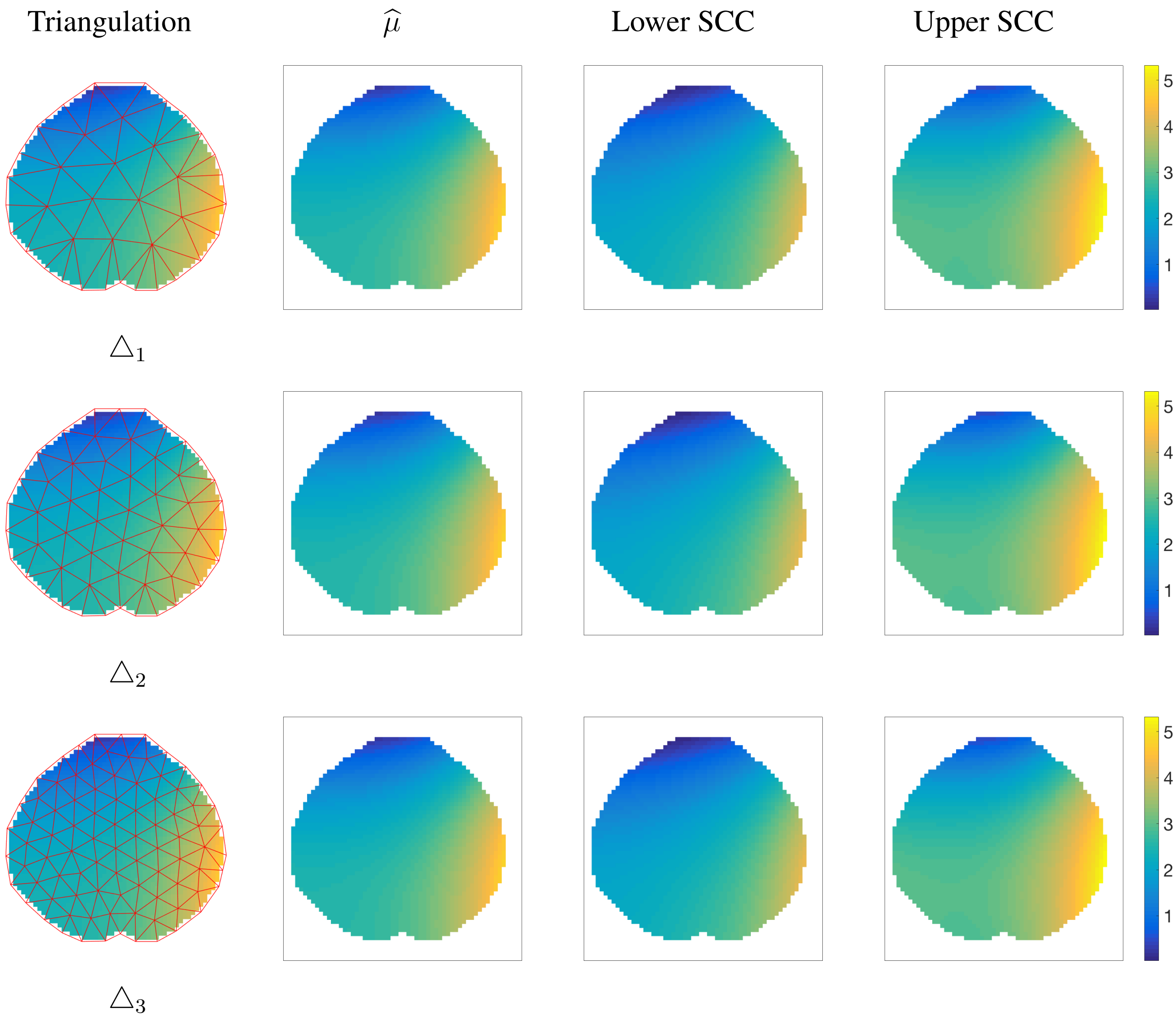


Figure A5: SCCs for cubic function with $n = 50$ and $\alpha = 0.01$.

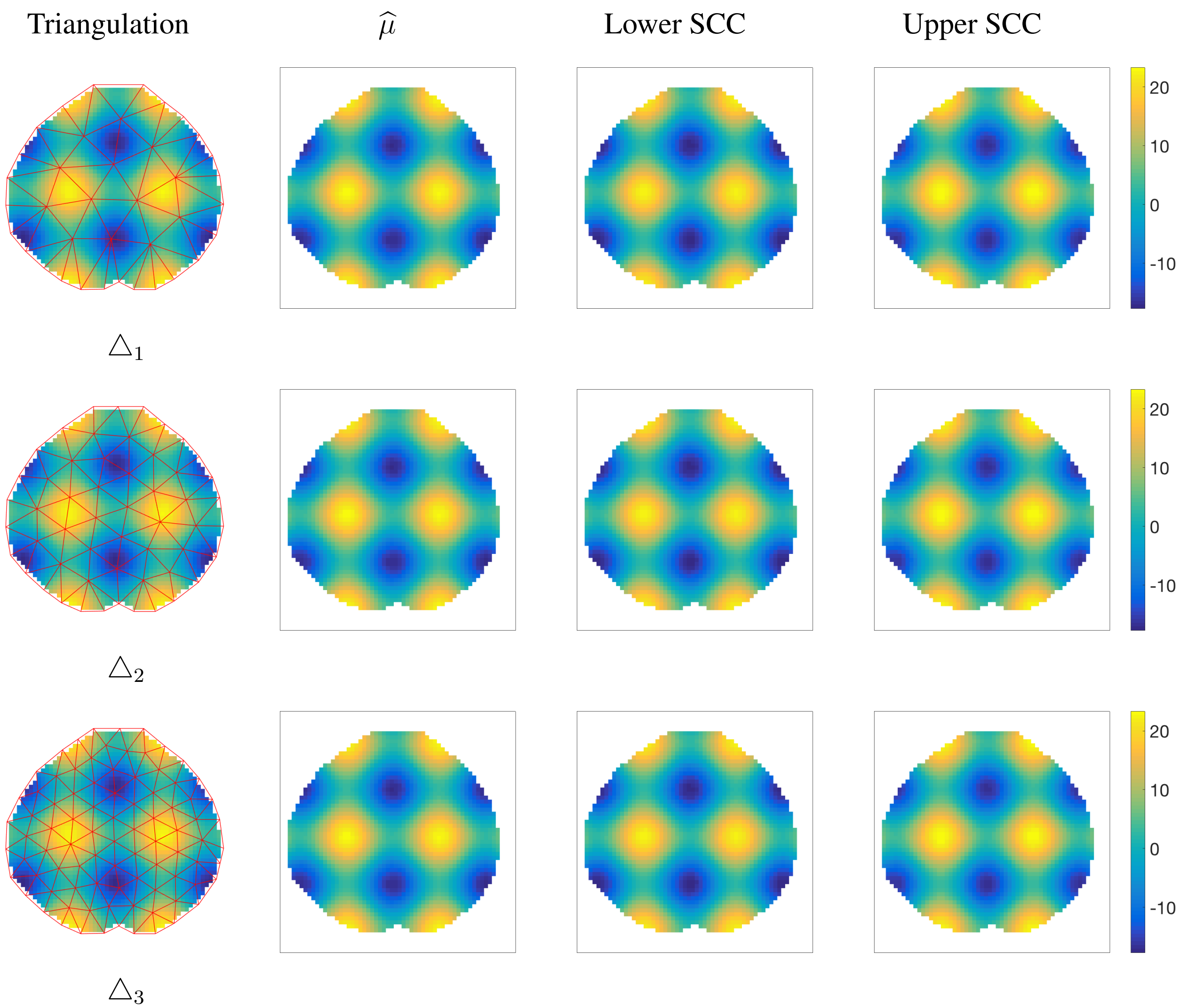


Figure A6: SCCs for sine function with $n = 50$ and $\alpha = 0.01$.

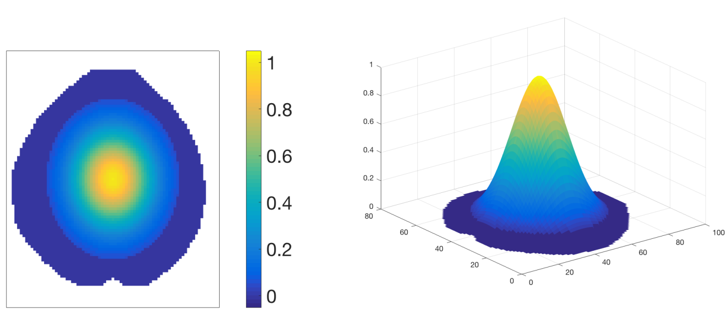
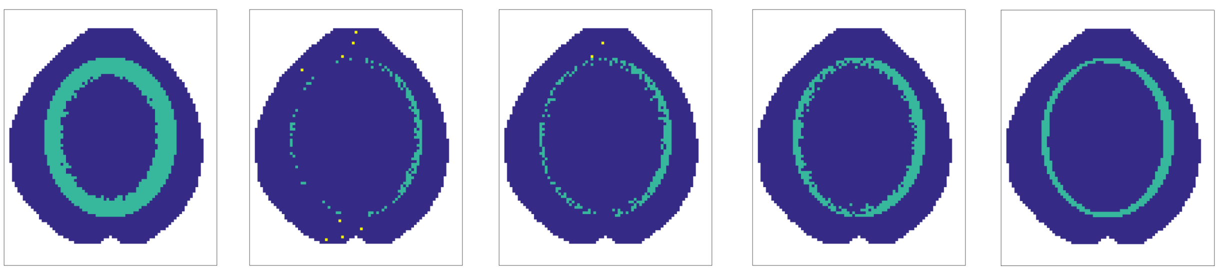


Figure A7: True mean function: (a) image map and (b) surface plot.

h



(a) Bonferroni (b) Cluster (0.1) (b) Cluster (0.05) (c) Cluster (0.01) (d) SCC

Figure A8: Signal discovery for one typical replication. Blue area shows the pixels correctly detected; yellow area shows the false positive pixels; and green area shows the false negative pixels.

Table A1: Empirical coverage rates of the SCCs ($N = 3682$).

n	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
	$\triangle_1$	$\triangle_2$	$\triangle_3$	$\triangle_1$	$\triangle_2$	$\triangle_3$	$\triangle_1$	$\triangle_2$	$\triangle_3$
$\mu(\boldsymbol{z}) = 20\left\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\right\}$									
50	0.871	0.876	0.876	0.937	0.938	0.937	0.982	0.984	0.984
	(0.643)	(0.644)	(0.644)	(0.731)	(0.732)	(0.733)	(0.902)	(0.903)	(0.903)
100	0.885	0.881	0.882	0.939	0.942	0.941	0.979	0.979	0.979
	(0.460)	(0.458)	(0.458)	(0.522)	(0.521)	(0.521)	(0.643)	(0.641)	(0.642)
200	0.901	0.902	0.883	0.949	0.949	0.941	0.987	0.988	0.987
	(0.330)	(0.331)	(0.326)	(0.374)	(0.375)	(0.370)	(0.460)	(0.461)	(0.457)
$\mu(\boldsymbol{z}) = 5\exp\left[-15\left\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\right\}\right] + 0.5$									
50	0.868	0.871	0.871	0.934	0.934	0.934	0.982	0.984	0.982
	(0.643)	(0.644)	(0.644)	(0.731)	(0.732)	(0.733)	(0.902)	(0.903)	(0.903)
100	0.896	0.893	0.880	0.945	0.944	0.938	0.980	0.981	0.979
	(0.465)	(0.464)	(0.458)	(0.527)	(0.526)	(0.521)	(0.648)	(0.647)	(0.642)
200	0.901	0.899	0.898	0.947	0.947	0.949	0.987	0.988	0.988
	(0.330)	(0.331)	(0.331)	(0.374)	(0.375)	(0.376)	(0.460)	(0.461)	(0.462)
$\mu(\boldsymbol{z}) = 3.2(-z_1^3 + z_2^3) + 2.4$									
50	0.860	0.870	0.869	0.927	0.931	0.929	0.985	0.987	0.987
	(0.628)	(0.628)	(0.629)	(0.716)	(0.716)	(0.718)	(0.887)	(0.887)	(0.889)
100	0.892	0.894	0.895	0.942	0.947	0.947	0.982	0.983	0.983
	(0.451)	(0.451)	(0.452)	(0.514)	(0.514)	(0.515)	(0.635)	(0.635)	(0.635)
200	0.899	0.902	0.898	0.942	0.947	0.949	0.988	0.988	0.989
	(0.320)	(0.320)	(0.320)	(0.364)	(0.365)	(0.365)	(0.451)	(0.451)	(0.451)
$\mu(\boldsymbol{z}) = -10[\sin\{5\pi(z_1 + 0.22)\} - \sin\{5\pi(z_2 - 0.18)\}] + 2.8$									
50	0.885	0.892	0.867	0.940	0.943	0.928	0.983	0.985	0.982
	(0.703)	(0.717)	(0.719)	(0.790)	(0.804)	(0.807)	(0.959)	(0.973)	(0.977)
100	0.894	0.890	0.883	0.943	0.946	0.934	0.979	0.981	0.977
	(0.500)	(0.509)	(0.516)	(0.562)	(0.571)	(0.578)	(0.681)	(0.691)	(0.699)
200	0.899	0.899	0.892	0.946	0.947	0.946	0.988	0.988	0.987
	(0.354)	(0.361)	(0.368)	(0.398)	(0.405)	(0.412)	(0.483)	(0.490)	(0.497)

Table A2: Type I error and empirical power of two-sample test

n	$\triangle$	δ 0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80
						$\alpha = 0.10$				
50	49	0.110	0.204	0.374	0.620	0.842	0.974	1.000	1.000	1.000
	80	0.102	0.194	0.366	0.612	0.844	0.972	1.000	1.000	1.000
	144	0.101	0.192	0.367	0.616	0.844	0.976	1.000	1.000	1.000
100	49	0.108	0.249	0.560	0.886	0.999	1.000	1.000	1.000	1.000
	80	0.106	0.242	0.549	0.884	1.000	1.000	1.000	1.000	1.000
	144	0.103	0.247	0.559	0.886	0.999	1.000	1.000	1.000	1.000
200	49	0.087	0.334	0.848	1.000	1.000	1.000	1.000	1.000	1.000
	80	0.085	0.319	0.836	1.000	1.000	1.000	1.000	1.000	1.000
	144	0.082	0.325	0.844	1.000	1.000	1.000	1.000	1.000	1.000
						$\alpha = 0.05$				
50	49	0.053	0.110	0.250	0.474	0.700	0.900	0.992	1.000	1.000
	80	0.049	0.101	0.244	0.467	0.692	0.894	0.988	1.000	1.000
	144	0.051	0.107	0.252	0.472	0.699	0.899	0.989	1.000	1.000
100	49	0.058	0.153	0.414	0.779	0.973	1.000	1.000	1.000	1.000
	80	0.056	0.150	0.405	0.766	0.966	1.000	1.000	1.000	1.000
	144	0.056	0.151	0.415	0.770	0.969	1.000	1.000	1.000	1.000
200	49	0.037	0.217	0.697	0.992	1.000	1.000	1.000	1.000	1.000
	80	0.037	0.211	0.685	0.992	1.000	1.000	1.000	1.000	1.000
	144	0.035	0.217	0.696	0.992	1.000	1.000	1.000	1.000	1.000
						$\alpha = 0.01$				
50	49	0.014	0.026	0.088	0.241	0.462	0.692	0.882	0.982	1.000
	80	0.012	0.025	0.087	0.228	0.453	0.677	0.875	0.977	1.000
	144	0.010	0.027	0.089	0.235	0.463	0.690	0.882	0.983	1.000
100	49	0.013	0.032	0.181	0.509	0.825	0.976	1.000	1.000	1.000
	80	0.012	0.032	0.172	0.486	0.817	0.978	0.999	1.000	1.000
	144	0.012	0.032	0.186	0.509	0.828	0.979	0.999	1.000	1.000
200	49	0.009	0.071	0.417	0.890	0.999	1.000	1.000	1.000	1.000
	80	0.009	0.065	0.402	0.884	0.998	1.000	1.000	1.000	1.000
	144	0.009	0.068	0.420	0.884	0.998	1.000	1.000	1.000	1.000

Table A3: FPRs, FNRs and FDRs for different methods.

n	Criterion	Method				
		Bonferroni	Cluster (0.10)	Cluster (0.05)	Cluster (0.01)	SCC
100	FPR	0.0000	0.0472	0.0233	0.0067	0.0090
	FNR	0.3158	0.1288	0.1567	0.2071	0.1868
	FDR	0.0000	0.0876	0.0449	0.0142	0.0169
200	FPR	0.0000	0.0534	0.0260	0.0044	0.0043
	FNR	0.2497	0.0836	0.1051	0.1485	0.1377
	FDR	0.0000	0.0893	0.0478	0.0081	0.0062